(54) Title: SALMONELLA VACCINES

(57) Abstract

A bacterial cell the virulence of which is attenuated by a first mutation in a PhoP regulon and a second mutation in an aromatic amino acid synthetic gene and bacterial cells the virulence of which is attenuated by a mutation in one or more PhoP-activated genes or one or more PhoP-repressed genes.

SALMONELLA VACCINES
Background of the Invention

5   The invention relates to vaccines.

           This invention was made with Government support
under Grant No. AI30479 and Grant No. 00917 awarded by
the National Institutes of Health. The Government has
certain rights in the invention.

10   Enteric fevers and diarrheal diseases, e.g.,
typhoid fever and cholera, are major causes of morbidity
and mortality throughout the developing world, Hook et
al., 1980, In Harrison's Principles of Internal Medicine,
9th Ed., 641-848, McGraw Hill, New York. Traditional
15 approaches to the development of vaccines for bacterial
diseases include the parenteral injection of purified
components or killed organisms. These parenterally
administered vaccines require technologically advanced
preparation, are relatively expensive, and are often,
20 because of dislike for needle-based injections, resisted
by patients. Live oral vaccine strains have several
advantages over parenteral vaccines: low cost, ease of
administration, and simple preparation.

           The development of live vaccines has often been
25 limited by a lack of understanding of the pathogenesis of
the disease of interest on a molecular level. Candidate
live vaccine strains require nonrevertable genetic
alterations that affect the virulence of the organism,
but not its induction of an immune response. Work
30 defining the mechanisms of toxigenesis of *Vibrio cholerae*
has made it possible to create live vaccine strains based
on deletion of the toxin genes, Mekalanos et al., 1983,
Nature 306:551, Levine et al., 1988, Infect. Immun.
56:161.

35   Recent studies have begun to define the molecular
basis of *Salmonella typhimurium* macrophage survival and

- 2 -

virulence, Miller et al., 1989, Proc. Natl. Acad. Sci.
USA 86:5054, hereby incorporated by reference.
Salmonella typhimurium strains with mutations in the
positive regulatory regulon *phoP* are markedly attenuated
5   in virulence for BALB/c mice. The *phoP* regulon is
composed of two genes present in an operon, termed *phoP*
and *phoQ*. The *phoP* and *phoQ* gene products are highly
similar to other members of bacterial two-component
transcriptional regulators that respond to environmental
10   stimuli and control the expression of a large number of
other genes. A mutation at one of these *phoP* regulatory
region regulated genes, *pagC*, confers a virulence defect.
Strains with *pagC*, *phoP*, or *phoQ* mutations afford partial
protection to subsequent challenge by wild-type *S.*
15   *typhimurium.*

        Salmonella species cause a spectrum of clinical
disease that includes enteric fevers and acute
gastroenteritis, Hook et al., 1980, *supra*. Infections
with *Salmonella* species are more common in
20   immunosuppressed persons, Celum et al., 1987, J. Infect.
Dis. 156:998. *S. typhi*, the bacterium that causes
typhoid fever, can only infect man, Hook et al., 1980,
supra. The narrow host specificity of *S. typhi* has
resulted in the extensive use of *S. enteriditis*
25   *typhimurium* infection of mice as a laboratory model of
typhoid fever, Carter et al., 1984 J. Exp. Med. 139:1189.
S. typhimurium infects a wider range of hosts, causing
acute gastroenteritis in man and a disease similar to
typhoid fever in the mouse and cow.

30   *Salmonella* infections are acquired by oral
ingestion. The organisms, after traversing the stomach,
replicate in the small bowel, Hornik et al., 1970, N.
Eng. J. Med. 283:686. *Salmonella* are capable of invasion
of the intestinal mucosal cells, and *S. typhi* can pass
35   through this mucosal barrier and spread via the Peyer's

- 3 -

patches to the lamina propria and regional lymph nodes.
Colonization of the reticuloendothelial cells of the host
then occurs after bacteremia. The ability of *S. typhi* to
survive and replicate within the cells of the human
5 reticuloendothelial system is essential to its
pathogenesis, Hook et al., 1980, *supra*, Hornick et al.,
1970, *supra*, and Carter et al., 1984, *supra*.

      Immunity to *Salmonella typhi* involves humoral and
cell-mediated immunity, Murphy et al., 1987, J. Infect.
10 Dis. 156:1005, and is obtainable by vaccination, Edelman
et al., 1986, Rev. Inf. Dis. 8:324. Recently, human
field trials demonstrated significant protective efficacy
against *S. typhi* infection after intramuscular
vaccination with partially purified Vi antigen, Lanata et
15 al., 1983, Lancet 2:441. Antibody-dependent enhancement
of *S. typhi* killing by T cells has been demonstrated in
individuals who received a live *S. typhi* vaccine,
indicating that these antibodies may be necessary for the
host to generate a cell-mediated immune response, Levine
20 et al., 1987, J. Clin. Invest. 79:888. The cell-mediated
immune response is important in typhoid immunity since
killed vaccines that do not induce this immune response
are not protective in man, Collins et al., 1972, Infect.
Immun. 41:742.

25   Summary of the Invention

      The invention provides a *Salmonella* vaccine which
does not cause transient bacteremia. In general, the
invention features a bacterial cell, preferably a
Salmonella cell, e.g., a *S. typhi*, *S. enteritidis*
30 *typhimurium*, or *S. cholerae-suis* cell, the virulence of
which is attenuated by a first mutation in a PhoP regulon
and a second mutation in an aromatic amino acid synthetic
gene. As used herein, PhoP regulon is defined as a DNA
which comprises a unit of *Salmonella* virulence gene
35 expression characterized by two regulatory genes, *phoP*

- 4 -

and *phoQ*, and structural genes, the expression of which
is regulated by *phoP* and *phoQ*, e.g., phoP regulatory
region repressed genes (prg) or phoP regulatory region
activated genes (pag). Such a bacterial cell can be used

5   as a vaccine to immunize a mammal against salmonellosis.

The *Salmonella* cell may be of any serotype, e.g.,
S. typhimurium, *S. paratyphi A*, *S. paratyphi B*, *S.
paratyphi C*, *S. pylorum*, *S. dublin*, *S. heidelberg*, *S.
newport*, *S. minnesota*, *S. infantis*, *S. virchow*, or *S.

10   panama*.

The first mutation may be a non-revertable null
mutation in the PhoP/PhoQ locus. Preferably, the
mutation is a deletion of at least 100 nucleotides; more
preferably, the mutation is a deletion of at least 500

15   nucleotides; even more preferably, the mutation is a
deletion of at least 750 nucleotides; and, most
preferably, the mutation is a deletion of nucleotides 376
to 1322 of the *PhoP/PhoQ* regulatory locus.

The second mutation may be a non-revertable null

20   mutation in an *aroA* locus or a non-revertable null
mutation in an *aroC/aroD* locus, or another locus involved
in the biosynthesis of aromatic amino acids.

To further attenuate the virulence of the
bacterial cell of the invention, the cell may contain yet

25   another mutation, e.g., a deletion, in a non-aromatic
amino acid synthetic gene, e.g., a mutation which renders
the cell auxotrophic for a non-aromatic amino acid, e.g.,
histidine. In preferred embodiments, the bacterial cell
of the invention is a *S. typhi* cell with the genotype

30   AroA⁻, His⁻, PhoP/PhoQ⁻, e.g., TyLH445.

The invention may also include a *Salmonella* cell,
the virulence of which is attenuated by the constitutive
expression of a gene under the control of a two-component
regulatory system. In preferred embodiments the

35   constitutive expression is the result of a mutation at a

- 5 -

component of the two-component regulatory system. In
preferred embodiments the bacterial cell includes a
second mutation which attenuates virulence.

In yet other preferred embodiments of the vaccine
the two-component regulatory system is the phoP
regulatory region, and the gene under the control of the
two-component system is a phoP regulatory region
regulated gene, e.g., a *prg* gene, e.g., *prgA*, *prgB*, *prgC*,
prgE, or *prgH*, or *pag* gene, e.g., *pagC*. In preferred
embodiments constitutive expression is the result of a
change or mutation, e.g., a deletion, (preferably a non-
revertible mutation) at the promoter of the regulated
gene or of the phoP regulatory region, e.g., a mutation
in the *phoQ* or the *phoP* gene, e.g., the PhoP^c mutation.

In another aspect, the invention features a
vaccine including a bacterial cell which is attenuated by
decrease of expression of a virulence gene under control
of a *phoP* regulatory region, e.g., a *prg* gene, e.g.,
prgA, *prgB*, *prgC*, *prgE*, or *prgH*.

In preferred embodiments of the vaccine the
Salmonella cell includes a first mutation, e.g., a
deletion, which attenuates virulence, e.g., a mutation in
a *phoP* regulatory region gene, e.g., a mutation in the
phoP or *phoQ* gene, e.g., PhoP^c, or a mutation in a phoP
regulatory region regulated gene, and a second mutation
which attenuates virulence, e.g., a mutation in an
aromatic amino acid synthetic gene, e.g., an *aro* gene, a
mutation in a phoP regulatory region regulated gene,
e.g., a mutation in a *prg* gene, e.g., *prgA*, *prgB*, *prgC*,
prgE, or *prgH*, or *pag* locus, e.g., a *pagC* mutation.

In yet other preferred embodiments the bacterial
cell includes a first mutation in a phoP regulatory
region gene and a second mutation in an *aromatic* amino
acid synthetic gene, e.g, an *aro* gene.

In another aspect, the invention features a

- 6 -

vaccine, preferably a live vaccine, including a bacterial
cell, the virulence of which is attenuated by a mutation,
e.g., a deletion, in a gene under the control of a two-
component regulatory system. In preferred embodiments

5   the bacterial cell includes a virulence attenuating
mutation in a second gene, e.g., in an *aromatic* amino
acid synthetic gene, e.g., an *aro* gene.

In yet other preferred embodiments of the vaccine
the bacterial cell is a *Salmonella* cell, the two-

10   component regulatory system is the phoP regulatory
region, and the gene under its control is a *prg* gene,
e.g. *prgA*, *prgB*, *prgC*, *prgE*, or *prgH*, or a *pag* gene,
e.g., the *pagC* gene.

In another aspect the invention features a

15   vaccine, preferably a live vaccine, including a
Salmonella cell e.g., a *S. typhi*, *S. enteritidis*
typhimurium, or *S. cholerae-suis* cell, including a first
virulence attenuating mutation in an *aromatic* amino acid
biosynthetic gene, e.g., an *aro* gene, and a second

20   virulence attenuating mutation in a phoP regulatory
region gene, e.g., a *phoP⁻* mutation.

In another aspect the invention features a
bacterial cell, or a substantially purified preparation
thereof, preferably a *Salmonella* cell, e.g., a *S. typhi,*

25   *S. enteritidis typhimurium*, or *S. cholerae-suis* cell,
which constitutively expresses a gene under the control
of a two-component regulatory system and which includes a
virulence attenuating mutation, e.g., a deletion, which
does not result in constitutive expression of a gene

30   under the control of the two-component regulatory system.
In preferred embodiments the bacterial cell includes a
mutation in a component of the two-component regulatory
system.

In preferred embodiments the bacterial cell is a

35   *Salmonella* cell which expresses a phoP regulatory region

- 7 -

regulated gene constitutively (the constitutive
expression preferably caused by a mutation, preferably a
non-revertible mutation, e.g., a deletion in the phoP
regulatory region, e.g., a mutation in the *phoQ* or *phoP*
5   gene, e.g., phoP^c), and which includes a virulence
attenuating mutation, preferably a non-revertible
mutation, e.g., a deletion, preferably in an *aromatic*
amino acid synthetic gene, e.g., an *aro* gene, or in a
phoP regulatory region regulated gene, e.g., a *prg* gene,
10   e.g., *prgA*, *prgB*, *prgC*, *prgE*, or *prgH* or *pag* gene, e.g.,
pagC which does not result in the constitutive expression
of a gene under the control of the phoP regulatory
region.

In another aspect, the invention features a
15   bacterial cell, or a substantially purified preparation
thereof, e.g., a *Salmonella* cell, e.g., a *S. typhi* cell,
an *S. enteritidis typhimurium* or a *S. cholerae-suis* cell,
including a virulence attenuating mutation in a gene
regulated by a two-component regulatory system. In
20   preferred embodiments the virulence attenuating mutation
is in a phoP regulatory region regulated gene, e.g., a
prg gene, e.g., *prgA*, *prgB*, *prgC*, *prgE*, or *prgH* or *pag*
gene, e.g., *pagC*.

In preferred embodiments the bacterial cell
25   includes a second mutation, e.g., in an *aromatic* amino
acid synthetic gene, e.g., an *aro* gene, in a phoP
regulatory region gene, e.g., the phoP or phoQ genes, or
in a phoP regulating region regulated gene, e.g., a *prg*
gene, e.g., *prgA*, *prgB*, *prgC*, *prgE*, or *prgH* or a *pag*
30   gene, e.g., *pagC*, which attenuates virulence but which
does not result in constitutive expression of a phoP
regulatory region regulated gene.

The invention also features a live *Salmonella*
cell, or a substantially purified preparation thereof,
35   e.g., a *S. typhi*, *S. enteriditis typhimurium*, or

- 8 -

S. cholerae-suis cell, in which there is inserted into a
virulence gene, e.g., a gene in the phoP regulating
region, or a phoP regulating region regulated gene, e.g.,
a *prg* gene, e.g., *prgA*, *prgB*, *prgC*, *prgE*, or *prgH* or a
5   *pag* locus, e.g., *pagC*, a gene encoding a heterologous
protein, or a regulatory element thereof.

       In preferred embodiments the live *Salmonella* cell
carries a second mutation, e.g., an *aro* mutation, e.g.,
an *aroA* mutation, e.g., *aroA⁻* or *aroADEL407*, that
10   attenuates virulence.

       In preferred embodiments the DNA encoding a
heterologous protein is under the control of an
environmentally regulated promoter. In other preferred
embodiments the live *Salmonella* cell further includes a
15   DNA sequence encoding T7 polymerase under the control of
an environmentally regulated promoter and a T7
transcriptionally sensitive promoter, the T7
transcriptionally sensitive promoter controlling the
expression of the heterologous antigen.

20   The invention also features a vector capable of
integrating into the chromosome of *Salmonella* including:
a first DNA sequence encoding a heterologous protein; a
second (optional) DNA sequence encoding a marker e.g., a
selective marker, e.g., a gene that confers resistance
25   for a heavy metal resistance or a gene that complements
an auxotrophic mutation carried by the strain to be
transformed; and a third DNA sequence, e.g., a *phoP*
regulon encoded gene, e.g., a *prg* gene, e.g., *prgA*, *prgB*,
prgC, *prgE*, or *prgH* or a *pag* locus, e.g., *pagC*, encoding
30   a *phoP* regulatory region regulated gene product necessary
for virulence, the third DNA sequence being mutationally
inactivated.

       In other preferred embodiments: the first DNA
sequence is disposed on the vector so as to mutationally
35   inactivate the third DNA sequence; the vector cannot

- 9 -

replicate in a wild-type *Salmonella* strain; the
heterologous protein is under the control of an
environmentally regulated promoter; and the vector
further includes a DNA sequence encoding T7 polymerase
5   under the control of an environmentally regulated
promoter and a T7 transcriptionally sensitive promoter,
the T7 transcriptionally sensitive promoter controlling
the expression of the heterologous antigen.

In another aspect the invention includes a method
10   of vaccinating an animal, e.g., a mammal, e.g., a human,
against a disease caused by a bacterium, e.g.,
Salmonella, including administering a vaccine of the
invention.

The invention also includes a vector including DNA
15   which encodes the *pagC* gene product; a cell transformed
with the vector; a method of producing the *pagC* gene
product including culturing the transformed cell and
purifying the *pagC* gene product from the cell or culture
medium; and a purified preparation of the *pagC* gene
20   product.

In another aspect the invention includes a method
of detecting the presence of *Salmonella* in a sample
including contacting the sample with *pagC* encoding DNA
and detecting the hybridization of the *pagC* encoding DNA
25   to nucleic acid in the sample.

The invention also includes a vector including DNA
which encodes the *prgH* gene product; a cell transformed
with the vector; a method of producing the *prgH* gene
product including culturing the transformed cell and
30   purifying the *prgH* gene product from the cell or culture
medium; and a purified preparation of the *prgH* gene
product.

In another aspect the invention includes a method
of detecting the presence of *Salmonella* in a sample
35   including contacting the sample with *prgH* encoding DNA

- 10 -

and detecting the hybridization of the *prgH* encoding DNA
to nucleic acid in the sample.

        In another aspect the invention features a method
of attenuating the virulence of a bacterium, the
bacterium including a two-component regulatory system,
including causing a gene under the control of the two-
component system to be expressed constitutively. In
preferred embodiments the bacterium is *Salmonella*, e.g.,
S. typhi, S. enteritidis typhimurium, or *S. cholerae-*
suis, and the two-component system is the phoP regulatory
region.

        In yet another aspect, the invention features a
substantially pure DNA which includes the sequence given
in SEQ ID NO:5 or a fragment thereof.

        The invention also includes a substantially pure
DNA which includes a sequence encoding *pagD*, e.g.,
nucleotides 91 to 354 of SEQ ID NO:5 (*pagD* open reading
frame (ORF)) and degenerate variants thereof that encode
a product with essentially the amino acid sequence given
in SEQ ID NO:6, as well as the *pagD* ORF and its 5' non-
coding region, nucleotides 4 to 814 of SEQ ID NO:15)
which contains the *pagD* promoter. DNA in the region
between the *pagC* ORF and the *pagD* ORF (nucleotides 4 to
814 of SEQ ID NO:15), DNA which includes the *pagC*
promoter (nucleotides 562 to 814 of SEQ ID NO:15), and
DNA which includes the *pagD* promoter alone (nucleotides 4
to 776 of SEQ ID NO:15) are also within the claimed
invention.

        The invention also includes a substantially pure
DNA which includes a sequence encoding *envE*, e.g.,
nucleotides 1114 to 1650 of SEQ ID NO:5 (*envE* ORF) and
degenerate variants thereof that encode a product with
essentially the amino acid sequence given in SEQ ID NO:7.

        Another aspect of the invention features a
substantially pure DNA which includes a sequence encoding

- 11 -

msgA, e.g., nucleotides 1825 to 2064 of SEQ ID NO:5 (*msgA* ORF) and degenerate variants thereof which encode a product with essentially the amino acid sequence given in SEQ ID NO:8, as well as the *msgA* ORF with its 5' non-
5   coding region, nucleotides 1510 to 1824 of SEQ ID NO:5 containing the *msgA* promoter. Also within the invention is a substantially pure DNA comprising the *msgA* promoter alone (nucleotides 1510 to 1760 of SEQ ID NO:5).

In yet another aspect, the invention features a
10 substantially pure DNA which includes a sequence encoding *envF*, e.g., nucleotides 2554 to 3294 of SEQ ID NO:5 (*envF* ORF) and degenerate variants thereof which encode a product with essentially the amino acid sequence given in SEQ ID NO:9, as well as the *envF* ORF with its 5' non-
15 coding region, nucleotides 2304 to 2553 of SEQ ID NO:5 which contains the *envF* promoter.

Also within the invention is a substantially pure DNA which includes the sequence given in SEQ ID NO:10 or a fragment thereof.

20   The invention also includes a substantially pure DNA which includes a sequence encoding *prgH*, e.g., nucleotides 688 to 1866 of SEQ ID NO:10 (*prgH* ORF) and degenerate variants thereof which encode a product with essentially the amino acid sequence given in SEQ ID
25 NO:11, as well as the *prgH* ORF with its promoter region (nucleotides 1 to 689 of SEQ ID NO:10).

The invention also includes a substantially pure DNA which includes a sequence encoding *prgI*, e.g., nucleotides 1891 to 2133 of SEQ ID NO:10 (*prgI* ORF) and
30 degenerate variants thereof which encode a product with essentially the amino acid sequence given in SEQ ID NO:12, as well as the *prgI* ORF with its promoter region (nucleotides 1 to 689 of SEQ ID NO:10.

In another aspect, the invention features a
35 substantially pure DNA which includes a sequence encoding

- 12 -

prgJ e.g., nucleotides 2152 to 2457 of SEQ ID NO:10 (*prgJ*
ORF) and degenerate variants thereof which encode a
product with essentially the amino acid sequence given in
SEQ ID NO:13, as well as the *prgJ* ORF and its promoter
5   region (nucleotides 1 to 689 of SEQ ID NO:10.

  In yet another aspect, the invention features a
substantially pure DNA which includes a sequence encoding
prgK, e.g., nucleotides 2456 to 3212 of SEQ ID NO:10
(*prgK* ORF) and degenerate variants thereof which encode a
10   product with essentially the amino acid sequence given in
SEQ ID NO:14, as well as the *prgK* ORF with its promoter
region (nucleotides 1 to 689 of SEQ ID NO:10.

  The invention also encompasses a bacterial cell
the virulence of which is attenuated by a mutation, e.g.,
15   a deletion, in one or more genes selected from the group
consisting of *pagD*, *pagE*, *pagF*, *pagG*, *pagH*, *pagI*, *pagJ*,
pagK, *pagL*, *pagM*, *pagN*, *pagP*, *envE*, and *envF*. Also
included is a bacterial cell which is attenuated by a
mutation, e.g, a deletion, in one or more genes selected
20   from the group consisting of *pagC*, *pagD*, *pagJ*, *pagK*,
pagM, and *msgA*. A bacterial cell, the virulence of
which is attenuated by a mutation, e.g., a deletion, in
one or more genes selected from the group consisting of
prgH, *prgI*, *prgJ*, and *prgK* is also within the claimed
25   invention.

  Two-component regulatory system, as used herein,
refers to a bacterial regulatory system that controls the
expression of multiple proteins in response to
environmental signals. The two-components referred to in
30   the term are a sensor, which may, e.g., sense an
environmental parameter and in response thereto promote
the activation, e.g. by promoting the phosphorylation, of
the second component, the activator. The activator
affects the expression of genes under the control of the
35   two-component system. A two-component system can

- 13 -

include, e.g., a histidine protein kinase and a
phosphorylated response regulator, as is seen in both
gram positive and gram negative bacteria. In *E. coli*,
e.g., 10 kinases and 11 response regulators have been
5   identified. They control chemotaxis, nitrogen
regulation, phosphate regulation, osmoregulation,
sporulation, and many other cellular functions, Stock et
al., 1989 Microbiol. Rev. 53:450-490, hereby incorporated
by reference. A two-component system also controls the
10   virulence of *Agrobacterium tumefasciens* plant tumor
formation, Leroux et al. EMBO J 6:849-856, hereby
incorporated by reference). Similar virulence regulators
are involved in the virulence of *Bordetella pertussis*
Arico et al., 1989, Proc. Natl. Acad. Sci. USA 86:6671-
15   6675, hereby incorporated by reference, and *Shigella
flexneri*, Bernardini et al., 1990, J. Bact. 172:6274-
6281, hereby incorporated by reference.
        Environmentally regulated, as used herein refers
to a pattern of expression wherein the expression of a
20   gene in a cell depends on the levels of some
characteristic or component of the environment in which
the cell resides. Examples include promoters in
biosynthetic pathways which are turned on or off by the
level of a specific component or components, e.g., iron,
25   temperature responsive promoters, or promoters which are
expressed more actively in specific cellular
compartments, e.g., in macrophages or vacuoles.
        A vaccine, as used herein, is a preparation
including materials that evoke a desired biological
30   response, e.g., an immune response, in combination with a
suitable carrier. The vaccine may include live organism,
in which case it is usually administered orally, or
killed organisms or components thereof, in which case it
is usually administered parenterally. The cells used for
35   the vaccine of the invention are preferably alive and

- 14 -

thus capable of colonizing the intestines of the
inoculated animal.

A mutation, as used herein, is any change (in
comparison with the appropriate parental strain) in the
5   DNA sequence of an organism. These changes can arise
e.g., spontaneously, by chemical, energy e.g., X-ray, or
other forms of mutagenesis, by genetic engineering, or as
a result of mating or other forms of exchange of genetic
information. Mutations include e.g., base changes,
10 deletions, insertions, inversions, translocations or
duplications.

A mutation attenuates virulence if, as a result of
the mutation, the level of virulence of the mutant cell
is decreased in comparison with the level in a cell of
15 the parental strain, as measured by (a) a significant
(e.g., at least 50%) decrease in virulence in the mutant
strain compared to the parental strain, or (b) a
significant (e.g., at least 50%) decrease in the amount
of the polypeptide identified as the virulence factor in
20 the mutant strain compared to the parental strain.

A non-revertible mutation, as used herein, is a
mutation which cannot revert by a single base pair
change, e.g., deletion or insertion mutations and
mutations that include more than one lesion, e.g., a
25 mutation composed of two separate point mutations.

The phoP regulatory region, as used herein, is a
two-component regulatory system that controls the
expression of *pag* and *prg* genes. It includes the phoP
locus and the phoQ locus.

30   phoP regulatory region regulated genes, as used
herein, refer to genes such as *pag* and *prg* genes.

pag, as used herein, refers to a gene which is
positively regulated by the phoP regulatory region.

prg, as used herein, refers to a gene which is
35 negatively regulated by the phoP regulatory region.

- 15 -

An *aromatic amino acid synthetic gene*, as used
herein, is a gene which encodes an enzyme which catalyzes
a step in the synthesis of an *aromatic amino acid*. *aroA*,
aroC, and *aroD* are examples of such genes in *Salmonella*.
5   Mutations in these genes can attenuate virulence without
the total loss of immunogenicity.

Abnormal expressions, as used herein, means
expression which is higher or lower than that seen in
wild type.

10 Heterologous protein, as used herein, is a protein
that in wild type, is not expressed or is expressed from
a different chromosomal site, e.g., a heterologous
protein is one encoded by a gene that has been inserted
into a second gene.

15 Virulence gene, as used herein, is a gene the
inactivation of which results in a *Salmonella* cell with
less virulence than that of a similar *Salmonella* cell in
which the gene is not inactivated. Examples include the
phoP, *pagC*, *prgH* genes.

20 A marker, as used herein, is gene product the
presence of which is easily determined, e.g., a gene
product that confers resistance to a heavy metal or a
gene product which allows or inhibits growth under a
given set of conditions.

25 Purified preparation, as used herein, is a
preparation, e.g., of a protein, which is purified from
the proteins, lipids, and other material with which it is
associated. The preparation is preferably at least 2-10
fold purified.

30 Constitutive expression, as used herein, refers to
gene expression which is modulated or regulated to a
lesser extent than the expression of the same gene in an
appropriate control strain, e.g., a parental or in wild-
type strain. For example, if a gene is normally
35 repressed under a first set of conditions and derepressed

- 16 -

under a second set of conditions constitutive expression
would be expression at the same level, e.g., the
repressed level, the derepressed level, or an
intermediate level, regardless of conditions. Partial
5   constitutive expression is included within the definition
of constitutive expression and occurs when the difference
between two levels of expression is reduced in comparison
in what is seen in an appropriate control strain, e.g., a
wild-type or parental strain.
10   A substantially purified preparation of a
bacterial cell is a preparation of cells wherein
contaminating cells without the desired mutant genotype
constitute less than 10%, preferably less than 1%, and
more preferably less than 0.1% of the total number of
15   cells in the preparation.
           The invention allows for the attenuation of
virulence of bacteria and of vaccines that include
bacteria, especially vaccines that include live bacteria,
by mutations in two-component regulatory systems and/or
20   in genes regulated by these systems. The vaccines of the
invention are highly attenuated for virulence but retain
immunogenicity, thus they are both safe and effective.
The vectors of the invention allow the rapid construction
of strains containing DNA encoding heterologous proteins,
25   e.g., antigens. The heterologous protein encoding DNA is
chromosomally integrated, and thus stable, unlike plasmid
systems which are dependent on antibiotic resistance or
other selection pressure for stability. Live *Salmonella*
cells of the invention in which the expression of
30   heterologous protein is under the control of an
environmentally responsive promoter do not express the
heterologous protein at times when such expression would
be undesirable e.g., during culture, vaccine preparation,
or storage, contributing to the viability of the cells,
35   but when administered to humans or animals, express large

- 17 -

amounts of the protein. This is desirable because high
expression of many heterologous proteins in *Salmonella*
can be associated with toxicity to the bacterium. The
use of only a single integrated copy of the DNA encoding

5   the heterologous protein also contributes to minimal
expression of the heterologous protein at times when
expression is not desired. In embodiments where a
virulence gene, e.g., the *pagC* gene or the *prgH* gene,
contains the site of integration for the DNA encoding the

10   heterologous protein the virulence of the organism is
attenuated.

A substantially pure DNA, as used herein, refers
to a nucleic acid sequence, segment, or fragment, which
has been purified from the sequences which flank it in a

15   naturally occurring state, e.g., a DNA which has been
removed from the sequences which are normally adjacent to
the fragment, e.g., the sequences adjacent to the
fragment in the genome in which it naturally occurs. The
term also applies to DNA which has been substantially

20   purified from other components which naturally accompany
the DNA, e.g., DNA which has been purified from proteins
which naturally accompany it in a cell.

Other features and advantages of the invention
will be apparent from the following description of the

25   preferred embodiments and from the claims.

              Description of the Preferred Embodiments
          The drawings will first be described.

Drawings

          Fig. 1 is a graph of the survival of *Salmonella*

30   strains within macrophages.

          Fig. 2 is a map of the restriction endonuclease
sites of the *pagC* locus.

          Fig. 3 is a map of the DNA sequence of the *pag* C
region (SEQ ID NO:1).

35   Fig. 4 is a map of the location of *prgH* within the

- 18 -

hil locus. The arrows indicate the direction of
orientation of the neomycin promoter of Tn5B50 insertions
within the *hil* locus and the direction of transcription
of the *prgH1*::TnphoA fusion protein. Restriction
5   endonuclease sites are represented by B, *Bam*H1; H,
*Hind*III; X, *Xho*I; S, *Sac*I; V, *Eco*RV.

        Fig. 5 is a DNA sequence from the *prgH* gene
(plasmid pIB01) (SEQ ID NO:3).

        Fig. 6 is a bar graph showing a comparison of the
10 sensitivity of wild type (ATCC 14028), PhoP-null mutant
(CS015), and *pag*::TnphoA mutant strains to NP-1 defensin.
The y-axis represents the Defensin Killing Index (DKI)
which is a measure of bacteria killed on exposure to NP-
1. The DKI is defined as the logarithmic function of the
15 ratio of control bacteria to surviving bacteria incubated
with NP-1 [DKI=log (CFU without NP-1/CFU with NP-1)].
The individual bars represent the mean and standard error
of five separate experiments. The x-axis indicates the
allele mutated. The mean DKI for each of the *pag*::TnphoA
20 strains tested was determined not be different from that
of wild type *Salmonella*. ($P<0.05$). In contrast, the *phoP*
mutant was significantly different ($P<0.0001$).

        Fig. 7 is a diagram showing a partial physical map
of the restriction endonuclease sites of the *pagC*
25 chromosomal region. The mouse 50% lethal doses (LD_{50}) for
strains with transposon insertions in *pagD*, *envE*, *msgA*,
and *pagC* are shown above each gene. Horizontal arrows
demonstrate the direction of transcription. Vertical
arrows denote TnphoA insertions and the hollow triangle
30 denotes a MudJ insertion. Below the chromosomal map is a
representation of the DNA insert in plasmid pCAA9, which
was mutagenized with TnphoA and MudJ. Letter
designations: A, *Acc*I; C, *Cla*I; E, *Eco*RI; H, *Hpa*I; P,
*Pst*I; and V, *Eco*RV.

- 19 -

Fig. 8 is a DNA sequence of the region upstream of *pagC* and the translation of each ORF. The *Hpa*I and *Cla*I sites at the beginning and end of the region are indicated. Shine-Delgarno regions are underlined and
5 stem loop structures (potential Rho-independent terminators) are indicated with a line below and above the sequence. Arrow heads denote the location of the representative transposon insertion within each gene. Horizontal arrows in the *pagD* and *msgA* promoter regions
10 mark the transcriptional start sites, and asterisks mark the -10 and -35 sequences. The consensus lipid attachment site in EnvF is enclosed in brackets. The *pagD* ORF begins at nucleotide 91 and ends at nucleotide 354 of SEQ ID NO:5; the *envE* ORF begins at nucleotide
15 1114 and ends at nucleotide 1650 of SEQ ID NO:5; the *msgA* ORF begins at nucleotid 1825 and ends at nucleotide 2064 of SEQ ID NO:5; and the *envF* ORF begins at nucleotide 2554 and ends at nucleotide 3294 of SEQ ID NO:5.

Fig. 9 is a DNA sequence containing the *prgH*,
20 *prgI*, *prgJ*, and *prgK* genes. The start codon (ATG) of each gene is underlined, and the stop codon is indicated with an asterisk. The *prgH* ORF begins at nucleotide 688 and ends at 1866 of SEQ ID NO:10; the *prgI* ORF begins at nucleotide 1891 and ends at nucleotide 2133 of SEQ ID
25 NO:10; the *prgJ* ORF begins at nucleotide 2152 and ends at nucleotide 2457 of SEQ ID NO:10; and the *prgK* ORF begins at nucleotide 2454 and ends at nucleotide 3212 of SEQ ID NO:10.

Fig. 10 is a line graph showing the growth rates
30 of the parent *Salmonella* strain (AroA-) and the vaccine strain (AroA-, PhoP-).

Fig. 11 is a bar graph showing defensin sensitivity of mouse vaccine strains (*S. typhimurium*).

- 20 -

Fig. 12 is a bar graph showing phoP activation as measured by LacZ activity using the PagB:LacZ recorder fusion construct.

Fig. 13 is a bar graph showing defensin
5   sensitivity of *S. typhi* vaccine strain TyLH445 compared to the AroA⁻ parent strain.

Fig. 14A is a graph showing the relative expression of constitutive expression (610 and 617) and phoP regulated (PagC and pagD) expression of AP fusion
10 proteins.

Fig. 14B is a graph showing the immune response to lipopolysaccharide (LPS).

Fig. 14C is a graph showing the immune response to the model heterologous antigen, AP.
15   Fig. 15 is a DNA sequence containing the *pagC-pagD* intergenic region. *pagC* translational start site (ATG on the opposite DNA strand) is underlined (nucleotides 1-3 of SEQ ID NO:15). The *pagC* transcriptional start (nucleotide 562) is indicated with an arrow pointing
20 left. The *pagD* translational start (ATG) is underlined (nucleotides 815-817 of SEQ ID NO:15). The *pagD* transcriptional start is indicated with an arrow pointing right (nucleotide 776).

Strain Deposit
25   Under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure, deposit of the following materials has been made with the American Type Culture Collection (ATCC) of Rockville, MD,
30 USA.

Applicant's assignee, Massachusetts General Hospital, represents that the ATCC is a depository affording permanence of the deposit and ready accessibility thereto by the public if a patent is
35 granted. All restrictions on the availability to the

- 21 -

public of the material so deposited will be irrevocably
removed upon the granting of a patent. The material will
be available during the pendency of the patent
application to one determined by the Commissioner to be
5   entitled thereto under 37 CFR 1.14 and 35 U.S.C. §122.
The deposited material will be maintained with all the
care necessary to keep it viable and uncontaminated for a
period of at least five years after the most recent
request for the furnishing of a sample of the deposited
10 plasmid, and in any case, for a period of at least thirty
(30) years after the date of deposit or for the
enforceable life of the patent, whichever period is
longer. Applicants' assignee acknowledges its duty to
replace the deposit should the depository be unable to
15 furnish a sample when requested due to the condition of
the deposit.

          PhoP^c strain CS022 (described below) has been
deposited with the American Type Culture Collection
(Rockville, MD) and has received ATCC designation 55130.

20   The plasmid, pIB01, containing the prgH gene has
been deposited on July 9, 1993 with the American Type
Culture Collection (Rockville, MD) and has received ATCC
designation ATCC 75496.

<u>Constitutive Expression of the PhoP Regulon Attenuates</u>
25 <u>Salmonella Virulence and Survival within Macrophages</u>

          The phoP constitutive allele (PhoP^c), pho-24,
results in derepression of pag loci. Using diethyl
sulfate mutagenesis of S. typhimurium LT-2, Ames and co-
workers isolated strain TA2367 pho-24 (all strains,
30 materials, and methods referred to in this section are
described below), which contained a phoP locus mutation
that resulted in constitutive production of acid
phosphatase in rich media, Kier et al., 1979, J.
Bacteriol. <u>138</u>:155, hereby incorporated by reference.
35 This phoP-regulated acid phosphatase is encoded by the
phoN gene, a pag locus, Kier et al., 1979, supra, Miller

- 22 -

et al., 1989, supra.   To analyze whether the *pho-24*
allele increased the expression of other *pag* loci the
effect of the *pho-24* allele on the expression of other
pag loci recently identified as transcriptional (e.g.,
5   *pag*A and *pag*B) and translational (e.g., *pag*C) fusion
proteins that required *phoP* and *phoQ* for expression,
Miller et al., 1989, supra, was determined.   *pag* gene
fusion strains, isogenic except for the *pho-24* allele,
were constructed and assayed for fusion protein activity.
10   PhoP[c] derivatives of the *pag*A::Mu dJ and *pag*B::Mu dJ
strains produced 480 and 980 U, respectively, of β-
galactosidase in rich medium, an increase of 9- to 10-
fold over values for the fusion strains with a wild-type
phoP locus, see Table 1.
15   The *pag*C::Tn*phoA* gene fusion produced 350 U of
AP, an increase of three- to fourfold over that produced
in strain CS119, which is isogenic except for the *pho-24*
mutation, Miller et al., 1989, supra.   These results
compare with a ninefold increase in the acid phosphatase
20   activity in strain CS022 on introduction of the *pho-24*
allele.   Therefore, these available assays for *pag* gene
expression document that the *pho-24* mutation causes
constitutive expression of *pag* loci other than *phoN*.

- 23 -

Table 1: Bacterial strains and properties

Strain	Genotype	Enzyme activity (U)[a]		Reference or source
10428	Wild type	180	(A)	ATCC; Miller et al., 1989, supra
TA2367	*pho-24*	1,925	(A)	Kier et al., 1974, supra
CS003	Δ*phoP* Δ*purB*	<10	(A)	Miller et al., 1989, supra
CS022	*pho-24*	1,750	(A)	This work
CS023	*pho-24 phoN2* *zxx*::6251Tn10d-Cam	25	(A)	This work
CS012	*pagA1*::MU dJ	45	(B)	Miller et al., 1989, supra
CS013	*pagB1*::MU dJ	120	(B)	Miller et al., 1989, supra
CS119	*pagC1*::Tn*phoA phoN2* *zxx*::6251Tn10d-Cam	85	(C)	Miller et al., 1989, supra
SC024	*pagA1*::Mu dJ *pho-24*	450	(B)	This work
SC025	*pagB1*::Mu dJ *pho-24*	980	(B)	This work
SC026	*pagC1*::Tn*phoApho-24phoN2* *zxx*::6251Tn10d-Cam	385	(B)	This work
CS015	*phoP102*::Tn10d-Cam	<10	(A)	Miller et al., 1989, supra
TT13208	*phoP105*::Tn10d	<10	(A)	--[b]

[a] A. Acid phosphatase; B, β-galactosidase; C, alkaline phosphatase (AP).

[b] Gift of Ning Zhu and John Roth.

- 24 -

Identifications of protein species that are repressed as well as activated in the PhoP^c mutant strain

Whole-cell proteins of strain CS022 were analyzed to estimate the number of protein species that could be potentially regulated by the PhoP regulon. Remarkably, analysis by one-dimensional polyacrylamide gel electrophoresis of the proteins produced by strains with the PhoP^c phenotype indicated that some protein species were decreased in expression when many presumptive *pag* gene products were fully induced by the *pho-24* mutation. The proteins decreased in the PhoP^c strain might represent products of genes that are repressed by the PhoP regulator. Genes encoding proteins decreased by the *pho-24* allele are designated *prg* loci, for *phoP*-repressed genes. Comparison of wild-type, PhoP$^-$, and PhoP^c mutant strain proteins shows that growth in LB medium at 37°C represents repressing conditions for *pag* gene products and derepressing conditions for *prg* gene products.

To estimate the total number of potentially PhoP-regulated gene products, the total cell proteins of wild-type and PhoP^c mutant strains grown in LB were analyzed by two-dimensional gel electrophoresis. At least 40 species underwent major fluctuation in expression in response to the *pho-24* mutation.

Virulence defects of the PhoP^c strain

Remarkably, strains with the single *pho-24* mutation were markedly attenuated for virulence in mice (Table 2). The number of PhoP^c organisms (2×10^5) that killed 50% of BALB/c mice challenged (LD_{50}) by the intraperitoneal (i.p.) route was near that (6×10^5) of PhoP$^-$ bacteria, Miller et al., 1989, supra. The PhoP^c strains had growth comparable to wild-type organisms in rich and minimal media. The PhoP^c mutants were also tested for alterations in lipopolysaccharide, which could explain the virulence defect observed. Strain CS022 had

Table 2
Virulence and protective efficacy of PhoP^c and PhoP$^-$ Salmonella strains

Immunizing dose	No. of initial survivors/total	No. of survivors/total after wild-type challenge dose of:			
		5×10^7	5×10^5	5×10^4	5×10^3
PhoP^c organisms					
15	13/13				
50	4/4		5/5	4/5	4/4
1.5×10^2	11/11		4/4	3/3	/4
5×10^2	16/16				
1.5×10^3	5/5		3/3	2/2	4/4
5×10^3	4/4				
1.5×10^4	5/5		3/3	2/2	4/4
5×10^4	19/23				
1.5×10^5	5/5		3/3	2/2	4/4
5×10^5	1/4				1/1
5×10^6	0/6				
3×10^9(*)	5/5	5/5			
3×10^{10}(*)	5/5	5/5			
1.5×10^{11}(*)	5/5	5/5			
PhoP$^-$ organisms					
6×10^3	36/36		0/12	0/12	0/12
6×10^4	36/36		0/12	0/12	3/12
6×10^5	19/36		0/6	0/6	4/7
5×10^{10}(*)	7/7	3/7			

(*) Organisms were administered by the oral route. In all other experiments, organisms were administered by i.p. challenge.

- 26 -

normal sensitivity to phage P22, normal group B
reactivity to antibody to O antigen, and a
lipopolysaccharide profile identical to that of the
parent strain, as determined by polyacrylamide gel
5   electrophoresis and staining.

        Since the TA2367 *pho-24* strain was constructed by
chemical mutagenesis and could have another linked
mutation responsible for its virulence defect revertants
of the PhoP^c were isolated to determine whether the *pho-*
10   *24* allele was responsible for the attenuation of
virulence observed. Phenotype PhoP^c revertants,
identified by the normal levels of acid phosphatase in
rich medium, were isolated among the bacteria recovered
from the livers of mice infected with strain CS022. Six
15   separate phenotypic revertants, designated CS122 to
CS128, were found to be fully virulent (LD$_{50}$ of less than
20 organisms for BALB/c mice). The locus responsible for
the reversion phenotype was mapped in all six revertants
tested for virulence by bacteriophage P22 cotransduction
20   and had linkage characteristics consistent with the *phoP*
locus (greater than 90% linkage to *purB*). These data
indicate that these reversion mutations are not
extragenic suppressors but are intragenic suppressors or
true revertants of the *pho-24* mutation. Thus, the
25   virulence defect of PhoP^c mutants is probably the result
of a single revertible mutation in the *phoP* locus and not
the result of a second unrelated mutation acquired during
mutagenesis.
Reversion frequency of the PhoP^c phenotype
30   The reversion frequency of the PhoP^c mutation *in
vivo* in mice was investigated to assess whether reversion
could reduce the LD$_{50}$ of this strain. The presence of the
revertants of strain CS022 was tested for by
administering 10^6, 10^4, and 10^2 challenge organisms to
35   each of eight animals by i.p. injection. On day 7, three

- 27 -

animals died that received 10^6 PhoP^c organisms. On that
day, the livers and spleens of all animals were harvested
and homogenized in saline. After appropriate dilution,
10% of the tissue was plated on LB plates containing the

5   chromogenic phosphatase substrate XP. Revertants were
identified by their lighter blue colonies compared with
PhoP^c bacteria and were confirmed by quantitative acid
phosphatase assays. An estimated 10^7, 10^5, and 10^3
organisms per organ were recovered from animals at each

10   of the three respective challenge doses. Revertants were
identified only at the highest dose and comprised 0.5 to
1%, or 10^5 organisms per organ, at the time of death. It
is likely that revertants are able to compete more
effectively for growth in these macrophage-containing

15   organs, since strain CS022 is deficient in survival
within macrophages (see below). However, revertants were
not identified if fewer than 10^5 organisms were
administered in the challenge dose, suggesting that the
reversion frequency must be approximately 10^{-5}. The

20   reversion rate of the PhoP^c phenotype for CS022 bacteria
grown in LB is in fact $6x10^{-4}$ when scored by the same
colony phenotypes. The percentage of revertants
recovered from animals near death suggests that pressure
is applied *in vivo* that selects for revertants of the

25   PhoP^c phenotype and implies that the virulence defect
observed could be much greater quantitatively for a
strain with a nonrevertible PhoP^c mutation.
The PhoP^c strain is deficient in survival within
macrophages

30   Because of the importance of survival within
macrophages to *Salmonella* virulence Fields et al., 1986,
Proc. Natl. Acad. Sci. USA 83:5189, hereby incorporated
by reference, PhoP^c bacteria were tested for this
property. Strain CS022 was defective in the ability to

35   grow and persist in macrophages as compared with wild-
type organisms (Fig. 1). In Fig. 1 the survival of

- 28 -

strain CS022 (PhoP^c) (triangles) in cultured macrophages
is compared with that of wild-type *S. typhimurium* ATCC
10428 (cicles). The experiment shown is a representative
one. The difference between the two strains at 4 and 24
5   hours is significant ($P < 0.05$). PhoP$^-$ bacteria seemed
to have a macrophage survival defect qualitatively
similar to that of PhoP^c bacteria but survived
consistently better by two- to threefold in side-by-side
experiments. The increased recovery of organisms that
10   reverted to PhoP^c phenotype in mouse organs rich in
macrophage content is consistent with the reduced
macrophage survival of PhoP^c mutants in vitro.
Use of the PhoP^c strain as a live vaccine
        It has been previously reported that PhoP$^-$ strains
15   are useful as live vaccines in protecting against mouse
typhoid, Miller et al., 1989, supra. The immunogenicity
of PhoP^c when used as live attenuated vaccines in mice
was compared with the of PhoP$^-$. This was done by
simultaneous determination of survival, after graded
20   challenge doses with the wild-type strain ATCC 10428, in
mice previously immunized with graded doses of the two
live vaccine strains. CS015 *phoP::Tn10*d-Cam and CS022
pho-24, as well as a saline control. The results
obtained (Table 2) suggest the following conclusions: (i)
25   small i.p. doses of the PhoP^c strain (e.g., 15 organisms)
effectively protect mice from challenge doses as large as
5×10^5 bacteria (a challenge dose that represents greater
than 10^4 i.p. LD$_{50}$s), (ii) large doses of PhoP^c organisms
given orally completely protect mice from an oral
30   challenge consisting of 5×10^7 wild-type bacteria (over
200 oral wild-type LD$_{50}$s) and (iii) by comparison, a large
dose of PhoP$^-$ organisms (5×10^5) does not provide similar
protection. The reversion of the PhoP^c mutation in vivo
somewhat complicates the analysis of the use of these
35   strains as vaccines, since revertants of the CS022 strain

- 29 -

(i.e., wild-type cells) could increase immunogenicity).
However, we were unable to identify revertants by
examining 10% of the available spleen and liver tissue
from those mice that received 10^4 or fewer organisms.

5   <u>Strains, Materials and Methods</u>
         The strains, materials, and methods used in the
PhoP regulon work described above are as follows.
         American Type Culture Collection (ATCC) strain
14028, a smooth virulent strain of *S. typhimurium*, was
10   the parent strain for all virulence studies. Strain
TT13208 was a gift from Nang Zhu and John Roth. Strain
TA2367 was a generous gift of Gigi Stortz and Bruce Ames,
Kier et al., 1979, supra. Bacteriophage P22HT *int* was
used in transductional crosses to construct strains
15   isogenic except for *phoP* locus mutations, Davis et al.,
1980, Advanced Bacterial Genetics, p. 78, 87. Cold
Spring Harbor Laboratory, Cold Spring Harbor, NY, hereby
incorporated by reference. Luria broth was used as rich
medium, and minimal medium was M9, Davis et al., 1980,
20   supra. The chromogenic phosphatase substrate 5-bromo-4-
chloro-3indolyl phosphate (XP) was used to qualitatively
access acid and AP production in solid media.
         Derivatives of *S. typhimurium* ATCC 10428 with the
pho-24 mutation were constructed by use of strain TA2367
25   as a donor of the *purB* gene in a P22 transductional cross
with strain CS003 Δ*phoP* Δ*purB*, Miller et al., 1989,
supra. Colonies were then selected for the ability to
grow on minimal medium. A transductant designated CS022
(phenotype PhoP^c) that synthesized 1,750 U of acid
30   phosphatase in rich medium (a ninefold increase over the
wild-type level in rich medium) was used in further
studies.
         Derivatives of strains CS022 and CS023 *pho-24*
phoN2 zxx::6251Tn*10*d-Cam, and acid phosphatase-negative
35   derivative of CS022, containing *pag* gene fusions were

- 30 -

constructed by bacteriophage P22 transductional crosses,
using selection of TnphoA- or Mu dJ-encoded kanamycin
resistance. Strains were checked for the intact pag gene
fusion by demonstration of appropriate loss of fusion

5   protein activity on introduction of a phoP105::Tn10d or
phoP102::Tn10d-Cam allele.

Assays of acid phosphatase, AP, and β-
galactosidase were performed as previously described,
Miller et al., 1989, supra and are reported in units as

10   defined in Miller, 1972, Experiments in molecular
genetics, p. 352-355, Cold Spring Harbor Laboratory, Cold
Spring Harbor, NY, hereby incorporated by reference.

In the mouse virulence and vaccination studies
bacteria grown overnight in Luria broth were washed and

15   diluted in normal saline. The wild-type parent strain of
CS022 (ATCC 10428) was used for all live vaccine
challenge studies. This strain has a 50% lethal dose
(LD_{50}) for naive adult BALB/c mice of less than 20
organisms when administered by intraperitoneal (i.p.)

20   injection and 5×10^4 when administered orally in $NaHCO_3$.
Mice were purchased from Charles River Breeding
Laboratories, Inc. (Wilmington, Mass.) and were 5 to 6
weeks of age at initial challenge. All i.p. inoculations
were performed as previously described, Miller et al.,

25   1989, supra. Oral challenge experiments were performed
with bacteria grown in LB broth and concentrated by
centrifugation. The bacteria were resuspended in 0.1 M
$NaHCO_3$ to neutralize stomach acid, and administered as a
0.5-ml bolus to animals under ether anesthesia. Colony

30   counts were performed to accurately access the number of
organisms administered. All challenge experiments were
performed 1 month after i.p. inoculation and 6 weeks
after oral challenge. Challenge inocula were
administered by the same route as vaccinations. The care

35   of all animals was under institutional guidelines as set

- 31 -

by the animal are committees at the Massachusetts General
Hospital and Harvard Medical School.

Protein electrophoresis was performed as follows.
One-dimensional protein gel electrophoresis was performed
5   by the method of Laemmli, 1970, Nature 227:680, hereby
incorporated by reference, on whole-cell protein extracts
of stationary-phase cells grown overnight in Luria broth.
The gels were fixed and stained with Coomassie brilliant
blue R250 in 10% acetic acid-10% methanol. Two-
10   dimensional protein gel electrophoresis was performed by
method of O'Farrell, 1975, J. Biol. Chem. 250:4007,
hereby incorporated by reference, on the same whole-cell
extracts. Isoelectric focusing using 1.5% pH 3.5 to 10
ampholines (LKB Instruments, Baltimore, Md.) was carried
15   out for 9,600 V h (700 V for 13 h 45 min). The final
tube gel pH gradient extended from pH 4.1 to pH 8.1 as
measured by a surface pH electrode (BioRad Laboratories,
Richmond, Calif.) and colored acetylated cytochrome pI
markers (Calbiochem-Behring, La Jolla, Calif.) run in an
20   adjacent tube. The slab gels were silver stained, Merril
et al., 1984, Methods Enzymol. 104:441, hereby
incorporated by reference.

In the macrophage survival assays experiments were
performed as previously described, Miller et al., 1989,
25   supra, by the method of Buchmeier et al., 1989, Infect.
Immun. 57:1, hereby incorporated by reference, as
modified from the method of Lissner et al, 1983, J.
Immunol. 131:3006, hereby incorporated by reference.
Stationary-phase cells were opsonized for 30 min in
30   normal mouse serum before exposure to the cultured bone
marrow-derived macrophages harvested from BALB/c mice.
One hour after infection, gentamicin sulfate (8 μg/ml)
was added to kill extracellular bacteria. All time
points were done in triplicate and repeated on three
35   separate occasions.

- 32 -

PhoP^C Mutant Strains Are More Effective as Live Vaccines

PhoP^C mutant *S. typhimurium* are very effective
when used as a live vaccine against mouse typhoid fever
and are superior to PhoP⁻ bacteria. As few a 15 PhoP^C
5   bacteria protect mice against 10^5 LD_{50} (lethal doses 50%)
of wild type organisms by the intraperitoneal route
(Table 3). This suggests that *pag* gene products are
important antigens for protective immunity against mouse
typhoid. Preliminary results have documented that
10 antigens recognized by serum of chronic typhoid carriers
recognizes some *phoP*-regulated gene products of *S. typhi*.
If protective antigens are only expressed within the
host, then dead vaccines only grown in rich media may not
induce an immune response against these proteins.
15   The use of different *S. typhimurium* dead vaccine
preparations containing different mutations in the *phoP*
regulon was evaluated. As can be seen in Table 3 no dead
cell preparations (even those containing mixtures of
PhoP⁻ and PhoP^C bacteria) are as effective vaccines as are
20 live bacteria. This suggests that there are other
properties of live vaccines that increase immunogenicity
or that important non-PhoP-regulated antigens are not in
these preparations. The only protection observed in any
animals studied was at the lowest challenge dose for
25 those immunized with PhoP^C bacteria. This further
suggests that *phoP* activated genes are important
protective antigens.

Table 3

Salmonella with phoP regulon mutations used as a dead vaccine

Vaccination Strain	phenotype	Challenge dose of wild type organisms	
		6×10^3	6×10^5
None			
ATCC10428	wild type	(3)	(5)
CS015	PhoP⁻	(8)	(9)
CS022	PhoP^c	(10)	(13)
CS022/CS015	PhoP⁻/PhoP^c	2/7(*)	(14)
		(8)	(13)

CS015 = phoP102::Tn10d-Cam
CS022 = pho-24

BALB/c mice were immunized twice, 7 days apart, with 5×10^8 formalin-killed bacteria. Three weeks after the second vaccination, mice were challenged with wild-type organisms at the two doses indicated. The numbers in parentheses indicate no survivors after challenge and mean number of days until death
(*) Ratio of survivors to number challenged.
   phoP^c indicates the constitutive unregulated expression of _phoP_-activated genes and lack of expression of _phoP_ repressed genes.
   phoP⁻ indicates a lack of expression of _phoP_-activated genes and expression of _phoP_ repressed genes.

- 34 -

aroA PhoP Regulon Double Mutant Strains

Recent efforts by Stocker, Levine, and colleagues have focused on the use of strains with auxotrophic mutations in _aromatic_ amino acid and purine pathways as
5   live vaccines, Hoseith et al., 1981, Nature 291:238, hereby incorporated by reference, Stocker, 1988, Vaccine 6:141, hereby incorporated by reference, and Levine et al., 1987, J. Clin. Invest. 79:888, hereby incorporated by reference. Purine mutations were found to be too
10   attenuating for immunogenicity, likely because purines are not available to the organism within the mammalian host, Sigwart et al., 1989, Infect. Immun. 57:1858, hereby incorporated by reference. Because auxotrophic mutations may be complemented by homologous recombination
15   events with wild type copies donated from environmental organisms or by acquiring the needed metabolite within the host, it would seem prudent for live vaccines to contain a second attenuating mutation in a different virulence mechanism, (i.e., not just a second mutation in
20   the same metabolic pathway). Additionally, in mice the _aroA_ mutants have some residual virulence. Various strains with _aroA_ mutations combined with _phoP_ regulon mutations were investigated for virulence attenuation and immunogenicity. Table 4 demonstrates that a PhoP$^-$ or
25   PhoP^c mutation further attenuates _aroA_ mutant _S. typhimurium_ by at least 100-fold and that, at least at high levels of vaccinating organisms, immunogenicity is retained. Strains with both a _pagC_$^-$ and _phoP_^{c} phenotype are also further attenuated than either mutation alone.
30   Therefore, _phoP_ regulon mutations may increase the safety of _aroA_ live vaccine preparations.

Table 4A

Additional attenuation of aroA mutants by PhoP regulon mutations

Survivors of varying numbers of
Salmonella mutant organisms (*)

Strain	Phenotype	10^6	10^7	10^8	10^9	$10^{10}(**)$
CS004	aroA-	6/6	1/6	0/6	0/6	6/6
SL3261	aroAdel His-	6/6	1/6	0/6	0/6	6/6
CS322	aroA- PhoP^c	6/6	6/6	6/6	1/6	6/6
CS323	SL3261 PhoP^c	6/6	6/6	6/6	2/6	6/6
CS315	aroA- PhoP-	6/6	6/6	6/6	2/6	6/6
CS316	SL3261 PhoP-	6/6	6/6	6/6	1/6	6/6
CS026	pagC$^-$ PhoP^c	6/6	4/6	0/6	0/6	6/6

Table 4B
Protective efficacy of _Salmonella_ with _aroA/phoP_ regulon mutations

Strain	Phenotype	Inoculum	Survivors of challenge doses of wild type organisms (*)	
			5×10^5	5×10^7
CS004	aroA⁻	10^6	4/4	5/5
SL3261	aroAdel His⁻	10^6	4/4	4/5
CS322	aroA⁻ PhoP^c	10^6	5/5	
CS323	SL3261 PhoP^c	10^6	5/5	
CS322	aroA⁻ PhoP^c	10^7	5/5	
CS323	SL3261 PhoP^c	10^7	5/5	
CS322	aroA⁻ PhoP^c	10^8		5/5
CS323	SL3261 PhoP^c	10^8		5/5
CS315	aroA⁻ PhoP⁻		5/5	
CS316	SI3261 PhoP⁻	10^8	5/5	

(*) Ratio of survivors to number of mice challenged.
(**) Indicates oral inoculation all other experiments were intraperitoneal inoculation.
CS004 = aroA554::rn10.
SL3261 = aroADEL407 hisG46.
CS322 = aroA554::Tn10 pho-24.
CS323 = aroADEL407 pho-24.
CS315 = aroA554::Tn10 phoP102::Tn10d-Cam.
CS316 = aroADEL407 hisG46 phoP102::Tn10d-Cam.
CS026 = pagC1::TnphoA pho-24 phoN2 zxx::6251TN10d-Cam.

- 37 -

Salmonella typhi phoP Regulon Mutations

The _phoP_ regulon is at least partially conserved
in _S. typhi_ DNA hybridization studies as well as P22
bacteriophage transductional crosses have documented that
5   the _phoP, phoQ_, and _pagC_ genes appear highly conserved
between _S. typhi_ and _S. typhimurium_ mutations in these
genes in _S. typhi_ have been made.

Salmonella Live Vaccines as Delivery Systems for
Heterologous Antigens

10   The vector used in the vaccine delivery system is
a derivative of pJM703.1 described in Miller et al.,
1988, J. Bact. <u>170</u>:2575, hereby incorporated by
reference. This vector is an R6K derivative with a
deletion in the _pir_ gene. R6K derivatives require the
15   protein product of the _pir_ gene to replicate. _E. coli_
that contain the _pir_ gene present as a lambda
bacteriophage prophage can support the replication of
this vector. Cells that do not contain the _pir_ gene will
not support the replication of the vector as a plasmid.
20   This vector also contains the _mob_ region of RP4 which
will allow mobilization into other gram negative bacteria
by mating from _E. coli_ strains such as SM10lambda _pir_,
which can provide the mobilization function in trans.

The _pagC_ region is shown in Figs. 2 and 3. Fig. 2
25   shows the restriction endonuclease sites of the _pagC_
locus. The heavy bar indicates _pagC_ coding sequence.
The Tn_phoA_ insertion is indicated by a inverted triangle.
The direction of transcription is indicated by the arrow
and is left to right. The numbers indicate the location
30   of endonuclease sites, in number of base pairs, relative
to the start codon of predicted _pagC_ translation with
positive numbers indicating location downstream of the
start codon and negative numbers indicating location
upstream of the start codon. A is _AccI_, B is _BglI_, C is
35   _ClaI_, D is _DraI_, E is _EcoRI_, H is _HpaI_, N is _NruI_, P is
PstI, S is _SspI_, T is _StuI_, U is _PvuII_, V is _EcoRV_, and

- 38 -

II is *Bgl*II.   Fig. 3 shows the DNA sequence (Sequence
I.D. No. 1) and translation of *pagC*::Tn*phoA*. The heavy
underlined sequence indicates a potential ribosomal
binding site. The single and double light underlines
5   indicate sequences in which primers were constructed
complementary to these nucleotides for primer extension
of RNA analysis. The asterisk indicates the approximate
start of transcription. The arrow indicates the
direction of transcription. The boxed sequences indicate
10   a region that may function in polymerase binding and
recognition. The inverted triangle is the site of the
sequenced Tn*phoA* insertion junction. The arrow indicates
a potential site for single sequence cleavage.

      3 kilobases of DNA containing the *pagC* gene (from
15   the *PstI* restriction endonuclease site 1500 nucleotides
5' to the start of *pagC* translation to the *EcoRI*
restriction endonuclease site 1585 nucleotides downstream
of *pagC* translation termination) were inserted into the
pJM703.1 derivative discussed above. The *pagC* sequence
20   from the *ClaI* restriction endonuclease site was deleted
(490 nucleotides) and replaced with a synthetic
oligonucleotide polylinker that creates unique
restriction endonuclease sites. DNA encoding one or more
heterologous proteins, e.g., an antigen, can be inserted
25   into this site. This creates a vector which allows the
insertion of multiple foreign genes into the DNA
surrounding *pagC*.

      The vector can be mobilized into *Salmonella* by
mating or any other delivery system, e.g., heat shock,
30   bacteriophage transduction or electroporation. Since it
can not replicate, the vector can only insert into
Salmonella by site specific recombination with the
homologous DNA on both sides of the *pagC* gene. This will
disrupt and inactivate the native *pagC* locus and replace
35   it with the disrupted *pagC* DNA carried on the vector.

- 39 -

Such recombination events can be identified by
marker exchange and selective media if the foreign DNA
inserted into the *pagC* locus confers a growth advantage.
The insertion of antibiotic resistance genes for
5   selection is less desirable as this could allow an
increase in antibiotic resistance in the natural
population of bacteria. Genes which confer resistance to
substances other than antibiotics e.g., to heavy metals
or arsenic (for mercury resistance, see Nucifora et al.,
10   1989, J. Bact., 171:4241-4247, hereby incorporated by
reference), can be used to identify transformants.
Alternatively, selection can be performed using a
Salmonella recipient strain that carries an auxotrophic
mutation in a metabolic pathway and a vector that carries
15   DNA that complements the auxotrophic mutation. Many
Salmonella live vaccine prototypes contain mutations in
histidine or purine pathways thus complementation of
these metabolic auxotrophies can be used to select for
integrants. (Purine mutations specifically have been
20   shown to be too attenuated for use in man.) Further
proof of marker exchange can be documented by loss of the
ampicillin resistance (carried on the plasmid backbone)
or by blot hybridization analysis.
A gene useful for selection can be cloned by
25   complementation of a vaccine strain with a metabolic
auxotrophy. Specific examples include the cloning of the
DNA encoding both *purB* and *phoP* by complementation of a
strain deleted for function of both these genes.
Salmonella gene libraries have been constructed in a
30   pLAFR cosmid vector (Frindberg et al., 1984, Anal.
Biochem. 137:266-267, hereby incorporated by reference)
by methods known to those skilled in the art. pLAFR
cosmids are broad host range plasmids which can be
mobilized into *Salmonella* from *E. coli*. An entire bank
35   of such strains can be mobilized into *Salmonella* vaccine

- 40 -

strains and selected for complementation of an
auxotrophic defect (e.g., in the case of *purB* growth on
media without adenine). The DNA able to complement this
defect is then identified and can be cloned into the
5   antigen delivery vector.

        As discussed above heterologous genes can be
inserted into the polylinker that is inserted into the
pagC sequence of the vector. The heterologous genes can
be under the control of any of numerous environmentally
10   regulated promotor systems which can be expressed in the
host and shut off in the laboratory. Because the
expression of foreign proteins, especially membrane
proteins (as are most important antigens), is frequently
toxic to the bacterium, the use of environmentally
15   regulated promoters that would be expressed in mammalian
tissues at high levels but which could be grown in the
laboratory without expression of heterologous antigens
would be very desirable. Additionally, high expression
of antigens in host tissues may result in increased
20   attenuation of the organism by diverting the metabolic
fuel of the organism to the synthesis of heterologous
proteins. If foreign antigens are specifically expressed
in host phagocytic cells this may increase the immune
response to these proteins as these are the cells
25   responsible for processing antigens.

        The promoter systems likely to be useful include
those nutritionally regulated promoter systems for which
it has been demonstrated that a specific nutrient is not
available to bacteria in mammalian hosts. Purines,
30   Sigwart et al., 1989, Infect. Immun., 57:1858 and iron,
Finklestein et al., 1983, Rev. Infect. Dis. 5:S759, e.g.,
are not available within the host. Promoters that are
iron regulated, such as the aerobactin gene promoter, as
well as promoters for biosynthetic genes in purine
35   pathways, are thus excellent candidates for testing as

- 41 -

promoters that can be shut down by growth in high
concentrations of these nutrients. Other useful
environmentally regulated *Salmonella* promoters include
promoters for genes which encode proteins which are
5   specifically expressed within macrophages, e.g., the DnaK
and GroEL proteins, which are increased by growth at high
temperature, as well as some *phoP* activated gene
products, Buchmeier et al., 1990, Science 248:730, hereby
incorporated by reference. Therefore, promoters such as
10   the *pagC* 5' controlling sequences and the better
characterized promoters for heat shock genes, e.g., GroEL
and DnaK, will be expected to be activated specifically
within the macrophage. The macrophage is the site of
antigen processing and the expression of heat shock genes
15   in macrophages and the wide conservation of heat shock
genes in nature may explain the immunodominance of these
proteins. A consensus heat shock promoter sequence is
known and can be used in the vectors (Cowling et al.,
1985, Proc. Natl. Acad. Sci. USA 82:2679, hereby
20   incorporated by reference).
         The vectors can include an environmentally
regulated T7 polymerase amplification system to express
heterologous proteins. For example, the T7 polymerase
gene (cloned by Stan Tabor and Charles Richardson, See
25   Current Protocols in Molecular Biology ed. Ausubel et
al., 1989, (page 3.5.1.2) John Wiley and Sons, hereby
incorporated by reference) under control of an iron
regulated promoter, can be included on the vectors
described above. We have inserted the aerobactin gene
30   promoter of *E. coli* with the sequence
CATTTCTCATTGATAATGAGAATCATTATTGACATAATTGTTATTATTTTACG
(SEQ ID NO:2), Delorenzo et al. J. Bact. 169:2624, hereby
incorporated by reference, in front of the T7 polymerase
gene and demonstrated iron regulation of the gene
35   product. This version of the vector will also include

- 42 -

one or more heterologous antigens under the control of T7
polymerase promoters. It is well known that RNA can be
synthesized from synthetic oligonucleotide T7 promoters
and purified T7 in vitro. When the organism encounters
5   low iron T7 polymerase will be synthesized and high
expression of genes with T7 promoters will be
facilitated.

pagC-fusion proteins in S. typhimurium

Expression of heterologous antigens within
10   macrophages under the control of _phoP_ regulated promotors
can be used as an effective method of both attenuating
Salmonellae and enhancing immunogenicity of foreign
antigens. As discussed above, the expression of PagC is
induced in antigen processing cell, i.e., a macrophage.
15   Thus, expression of a heterologous antigen under the
control of the _pagC_ promoter is also likely to be
inducible in macrophages.

To evaluate the immune respone to a heterologous
antigen expressed under the control of inducible _pag_
20   promoters, mice were inoculated with bacteria which
expressed the antigen, AP, under the control of the _pagC_
or _pagD_ regulatory sequences. Pag-AP fusion proteins
were produced in these bacteria from a single chromosomal
copy of the gene encoding AP. The bacteria were
25   generated utilizing two methods: Tn_phoA_ mutagenesis, and
genetic engineering techniques using a suicide vector,
both of which are described above.

As a control, mice were innoculated with bacteria
which expressed an AP fusion protein under the control of
30   constitutive promoters. The constitutive promoter was
completely independent of regulation by genes in the PhoP
regulon. Two such strains of bacteria, Strain 610 and
Strain 617, were constructed using methods described
above. AP expression in Strain 610 was moderate, whereas
35   AP expression in Strain 617 was high (see Fig. 14C).

- 43 -

        These strains were injected intraperitoneally into
BABL/C mice.   Serum samples were taken three weeks after
inoculation.   Normal mouse serum (MNS) was used as a
control.   Standard ELISA assays were used to test the
5 sera for the presence of AP-specific antibodies.   Sera
was also tested for LPS-specific antibodies using *S.
typhimurium* LPS.   Antibodies directed to LPS were
detected in all the murine sera tested, but only those
strains in which AP was expressed as a Pag fusion protein
10 from a single chromosomal gene copy engendered an immune
response against the model heterologous antigen, AP (see
Figs. 14A and Fig. 14B).

        Despite approximately 10-fold higher constitutive
expression of the AP fusion in strain 617, only a minimal
15 immune response to this antigen was noted after
immunization with strain 617.   In contrast, a strong
response was observed in mice inoculated with strains
which expressed the Pag-AP fusion protein.   These data
indicate that *phoP*-regulation which results in *in vivo*
20 induction of protein expression within macrophages
increases the immunogenicity of heterologous antigens
expressed under the control of the *pag* promoters.   Any
promoter which directs cell-specific, inducible
expression of a protein in macrophages or other antigen
25 presenting cells, e.g., *pag* described herein, can be used
to increase the immunogencity of an antigen expressed in
Salmonella.

The *pagC* gene and *pagC* Gene Product
Strains, materials, and methods
30   The following strains, materials, and methods were
used in the cloning of *pagC* and in the analysis of the
gene and its gene product.

        Rich media was Luria broth (LB) and minimal media
was M9, Davis et al., 1980, supra.   The construction of
35 *S. typhimurium* strain CS119 *pagC1*::Tn*phoA phoN2* zxx::6251
Tn*10d-Cam* was previously described, Miller et al., 1989,

- 44 -

supra. American Type Culture Collection (ATCC) *S. typhimurium* strain 10428 included CS018 which is isogenic to CS119 except for *phoP105*::Tn*10d*, Miller et al., 1989, supra, CS022 *pho-24*, Miller et al., 1990, J. Bacteriol.

5   $\underline{172}$:2485-2490, hereby incorporated by reference, and CS015 *phoP102*::Tn10d-cam, Miller et al., 1989, supra. Other wild type strains used for preparation of chromosomal DNA included *S. typhimurium* LT2 (ATCC 15277), *S. typhimurium* Q1 and *S. drypool* (Dr. J. Peterson U.

10 Texas Medical Branch, Galveston), and *Salmonella typhi* Ty2 (Dr. Caroline Hardegree, Food and Drug Administration). pLAFR cosmids were mobilized from *E. coli* to *S. typhimurium* using the *E. coli* strain MM294 containing pRK2013, Friedman et al., 1982, Gene $\underline{18}$:289-

15 296, hereby incorporated by reference. AP activity was screened on solid media using the chromogenic phosphatase substrate 5-bromo-4-chloro-3-indolyl phosphate (XP). AP assays were performed as previously described, Brickman et al., 1975, J. Mol. Biol. $\underline{96}$:307-316, hereby

20 incorporated by reference, and are reported in units as defined by Miller, Miller, 1972, supra, pp. 352-355.

One dimensional protein gel electrophoresis was performed by the method of Laemmli, 1970, Nature, $\underline{227}$:680-685, hereby incorporated by reference, and blot

25 hybridization using antibody to AP was performed as previously described, Peterson et al., 1988, Infect. Immun. $\underline{56}$:2822-2829, hereby incorporated by reference. Whole cell protein extracts were prepared, from saturated cultures grown in LB at 37°C with aeration, by boiling

30 the cells in SDS-*pag*E sample buffer, Laemmli, 1970, supra. Two dimensional gel electrophoresis was performed by the method of O'Farrell, 1975, J. Biol. Chem. $\underline{250}$:4007, hereby incorporated by reference. Proteins in the 10% polyacrylamide slab gels were visualized by

35 silver staining, Merril et al., 1984, Methods in

- 45 -

Enzymology, 104:441, hereby incorporated by reference.
       Chromosomal DNA was prepared by the method of
Mekalanos, 1983, Cell, 35:253-263, hereby incorporated by
reference. DNA, size fractionated in agarose gels, was
5 transferred to nitrocellulose (for blot hybridization) by
the method of Southern, 1975, J. Mol. Biol. 98:503-517,
hereby incorporated by reference. DNA probes for
Southern hybridization analysis were radiolabeled by the
random primer method, Frinberg et al., 1984, supra.
10 Plasmid DNA was transformed into E. coli and Salmonella
by calcium chloride and heart shock, Mekalanos, 1983,
supra, or by electroporation using a Genepulser apparatus
(Biorad, Richmond, Ca.) as recommended by the
manufacturer, Dower et al., 1988, Nucl. Acids Res.
15 16:6127-6145, hereby incorporated by reference. DNA
sequencing was performed by the dideoxy chain termination
method of Sanger et al., 1977, Proc. Natl. Acad. Sci.
USA, 74:5463-5467, hereby incorporated by reference, as
modified for use with SEQUENASE® (U.S. Biochemical,
20 Cleveland, Ohio). Oligonucleotides were synthesized on
an Applied Biosystems Machine and used as primers for
sequencing reactions and primer extension of RNA.
Specific primers unique to the two ends of TnphoA, one of
which corresponds to the AP coding sequence and the other
25 to the right IS50 sequence, were used to sequence the
junctions of the transposon insertion.
       Construction of a S. typhimurium cosmid gene bank
in pLAFR3 and screening for clones containing the wild
type pagC DNA was performed as follows. DNA from S.
30 typhimurium strain ATCC 10428 was partially digested
using the restriction endonuclease Sau3A and then size
selected on 10-40% sucrose density gradient. T4 DNA
ligase was used to ligate chromosomal DNA of size 20-30
kilobases into the cosmid vector pLAFR3, a derivative of
35 pLAFR1, Friedman et al., 1982, Gene 18:289-296, hereby

- 46 -

incorporated by reference, that was digested with the
restriction endonuclease *Bam*HI.   Cosmid DNA was packaged
and transfected into *E. coli* strain DH5-α using extracts
purchased from Stratagene, La Jolla, Ca.   Colonies were
5   screened by blot hybridization analysis.

The analysis of proteins produced from cloned DNA
by in vitro transcription/translation assays was analyzed
as follows.   These assays were performed with cell free
extracts, (Amersham, Arlington Heights, Illinois), and
10   were performed using conditions as described by the
manufacturer.   The resultant radiolabeled proteins were
analyzed by SDS-*pag*E.

RNA was purified from early log and stationary
phase *Salmonella* cultures by the hot phenol method, Case
15   et al., 1988, Gene 72:219-236, hereby incorporated by
reference, and run in agarose-formaldehyde gels for blot
hybridization analysis, Thomas, 1980, Proc. Natl. Acad.
Sci. USA 77:5201, hereby incorporated by reference.
Primer extension analysis of RNA was performed as
20   previously described, Miller et al., 1986, Nuc. Acids.
Res. 14:7341-7360, hereby incorporated by reference,
using AMV reverse transcriptase (Promega, Madison,
Wisconsin) and synthesized oligonucleotide primers
complementary to nucleotides 335-350 and 550-565 of the
25   *pag*C locus.

Identification of an 18 kDa protein missing in a *pag*C
mutant of *S. typhimurium*

*pag*C mutant strain CS119 was analyzed by two
dimensional protein electrophoresis to detect protein
30   species that might be absent as a result of the Tn*phoA*
insertion.   Only a single missing protein species, of
approximately 18 kD and pI-8.0, was observed when
strains, isogenic except for their transposon insertions,
were subjected to this analysis.   This 18 kDa species was
35   also missing in similar analysis of *Salmonella* strains
with mutations *phoP* and *phoQ*.   Though two-dimensional

- 47 -

protein gel analysis might not detect subtle changes of
protein expression in strain CS119, this suggested that a
single major protein species was absent as a result of
the pagC::TnphoA insertion.

5   Additional examination of the 2-dimensional gel
analysis revealed a new protein species of about 45 kDa
that is likely the pagC-Ap fusion protein. The pagC-AP
fusion protein was also analyzed by Western blot analysis
using antisera to AP and found to be similar in size to
10   native AP (45 kDa) and not expressed in PhoP-S.
typhimurium.
Cloning of the pagC::TnphoA insertion
        Chromosomal DNA was prepared from S. typhimurium
strain CS119 and a rough physical map of the restriction
15   endonuclease sites in the region of the pagC::TnphoA
fusion was determined by using a DNA fragment of TnphoA
as a probe in blot hybridization analysis. This work
indicated that digestion with the restriction
endonuclease ecoRV yielded a single DNA fragment that
20   included the pagC::TnphoA insertion in addition to
several kilobases of flanking DNA. Chromosomal DNA from
strain CS119 was digested with EcoRV (blunt end) and
ligated into the bacterial plasmid vector pUC19 (New
England Biolabs) that had been digested with the
25   restriction endonuclease SmaI (blunt end). This DNA was
electroporated into the E. coli strain DH5-α (BRL) and
colonies were plated onto LB agar containing the
antibiotics kanamycin (TnphoA encoded and ampicillin
(pUC19 encoded). A single ampicillin and kanamycin
30   resistant clone containing a plasmid designated pSM100
was selected for further study.
        A radiolabeled DNA probe from pSM100 was
constructed and used in Southern hybridization analysis
of strain CS119 and its wild type parent ATCC 10428 to
35   prove that the pagC::TnphoA fusion had been cloned. The

- 48 -

probe contained sequences immediately adjacent to the
transposon at the opposite end of the AP gene [*Hpa*I
endonuclease generated DNA fragment that included 186
bases of the right IS50 of the transposon and 1278 bases
5   of *Salmonella* DNA (Fig. 2). As expected, the pSM100
derived probe hybridized to an 11-12 kb *Acc*I endonuclease
digested DNA fragment from the strain containing the
transposon insertion, CS119. This was approximately
7.7kb (size of Tn*phoA*) larger than the 3.9 kB *Acc*I
10   fragment present in the wild type strain that hybridizes
to the probe. In addition, a derivative of plasmid
pSM100, pSM101 (which did not allow expression of the
pagC-PhoA gene fusion off the *lac* promoter), was
transformed into *phoP*- (strain Cs015) and *phoN*- (strain
15   CS019) *Salmonella* strains and the cloned AP activity was
found to be dependent on *phoP* for expression. Therefore
we concluded that the cloned DNA contained the
pagC::Tn*phoA* fusion.
         The presence of the *pagC* gene was also
20   demonstrated in other strains of *S. typhimurium*, as well
as in *S. typhi*, and *S. drypool*. All *Salmonella* strains
examined demonstrated similar strong hybridization to an
8.0 kb *Eco*RV and a 3.9 kb *Acc*I restriction endonuclease
fragment suggesting that *pagC* is a virulence gene common
25   to *Salmonella* species.
         The *pagC* gene probe from nucleotides -46 (with 1
as the first base of the methionine to 802 (*Pst*I site to
the *Bgl*II site) failed to cross hybridize to DNA from
Citrobacter freundii, *Shigella flexneri*, *Shigella sonnei*,
30   *Shigella dysenterial*, *Escherichia coli*, *Vibrio cholerae*,
Vibrio vulnificus, *Yersenia entero colitica*, and
Klebsiella pneumonia.

- 49 -

Cloning of the wild type pagC locus DNA and its complementation of the virulence defect of a S. typhimurium pagC mutant

The same restriction endonuclease fragment
5   described above was used to screen a cosmid gene bank of
wild type strain ATCC 10428. A single clone, designated
pWP061, contained 18 kilobases of S. typhimurium DNA and
hybridized strongly to the pagC DNA probe. pWP061 was
found to contain Salmonella DNA identical to that of
10 pSM100 when analyzed by restriction endonuclease analysis
and DNA blot hybridization studies. Probes derived from
pWP061 were also used in blot hybridization analysis with
DNA from wild type and CS119 S. typhimurium. Identical
hybridization patterns were observed to those seen with
15 pSM100. pWP061 was also mobilized into strain CS119, a
pagC mutant strain. The resulting strain had wild type
virulence for BALB/c mice (a LD$_{50}$ less than 20 organisms
when administered by IP injection). Therefore the cloned
DNA complements the virulence defect of a pagC mutant
20 strain.

Since, a wild type cosmid containing pagC locus
DNA was found to complement the virulence defect of a
pagC mutant S. typhimurium strain, it was concluded that
the pagC protein is an 188 amino acid (18 kDa) membrane
25 (see below) protein essential for survival within
macrophages and virulence of S. typhimurium.

Physical mapping of restriction endonuclease sites, DNA sequencing, and determination of the pagC gene product

Restriction endonuclease analysis of plasmid
30 pSM100 and pWP061 was performed to obtain a physical map
of the pagC locus, and, in the case of PSM100, to
determine the direction of transcription (Fig. 2). DNA
subclones were generated and the TnphoA fusion junctions
were sequenced, as well as the Salmonella DNA extending
35 from the HpaI site, 828 nucleotides 5' to the phoA fusion
junction, to the EcoRI site 1032 nucleotides 3' to the

- 50 -

TnphoA insertion (Fig. 2 and 3). The correct reading
frame of the DNA sequence was deduced from that required
to synthesize an active AP gene fusion. The deduced
amino acid sequence of this open reading frame was
5   predicted to encode a 188 amino acid protein with a
predicted pI+8.2. This data were consistent with the 2-D
polyacrylamide gel analysis of strain CS119 in which an
18 kDa protein of approximate pI+8.0 was absent. No
other open reading frames, predicted to encode peptides
10   larger than 30 amino acids, were found.

The deduced amino acid sequence of the 188 amino
acid open reading frame contains a methionine start codon
33 amino acids from the fusion of pagC and AP (Fig. 3).
This 33 amino acid pagC contribution to the fusion
15   protein was consistent with the size observed in Western
blot analysis and contains a hydrophobic N-terminal
region, identified by the method of Kyle et al., 1982, J.
Mol. Biol. 157:105-132, hereby incorporated by reference,
that is a typical bacterial signal sequence, Von Heinje,
20   1985, J. Mol. Biol. 184:99-105, hereby incorporated by
reference. Specifically, amino acid 2 is a positively
charged lysine, followed by a hydrophobic domain and
amino acid 24 is a negatively charged aspartate residue.
A consensus cleavage site for this leader peptide is
25   predicted to be at an alanine residue at amino acid 23,
Von Heinje, 1984, J. Mol. Biol. 173:243-251, hereby
incorporated by reference. The DNA sequence also
revealed a typical ribosomal binding site, Shine et al.,
1974, Proc. Natl. Acad. Sci. USA 71:1342-1346, hereby
30   incorporated by reference, at 6-2 nucleotides 5' to the
predicted start of translation (Fig. 3) nucleotides 717-
723). This suggested that the open reading frame was, in
fact, translated and further supported the assumption
that this was the deduced amino acid sequence of the pagC
35   protein interrupted by the TnphoA insertion (Fig. 3).

- 51 -

In vitro synthesis of proteins by the cloned pagC locus

        To detect if other proteins were encoded by pagC
and to determine the approximate size of the pagC gene
product, an in vitro coupled transcription/translation
analysis was performed. A 5.3 kilobase EcoRI fragment of
pWP061 was inserted into pUC19 so that the pagC gene
would not be expressed off the lac promotor. This
plasmid was used in an in vitro coupled transcription-
translation assay. A single protein of approximately 22
kilodaltons was synthesized by the cell free system. The
size was compatible with this being the precursor of the
pagC protein containing its leader peptide. These data
further support the conclusion the single and the single
pagC gene product had been identified.

Identification of the pagC encoded RNA

        An approximately 1100 nucleotide RNA is encoded by
pagC. The pagC gene is highly expressed by cells with a
phoP constitutive phenotype of pag activation, as
compared to wild type and phoP constitutive phenotype of
pag activation, as compared to wild type and phoP⁻
bacteria. In these blot hybridization experiments pagC
is only detected in wild type cells grown in rich media
during stationary growth. This result, coupled with
previous work, Miller et al., 1989, supra, Miller et al.,
1990, supra, demonstrates that pagC is transcriptionally
regulated by the phoP gene products and is only expressed
during early logarithmic phase growth in rich media by
cells with a phoP constitutive phenotype.

        The size of the pagC transcript is approximately
500 nucleotides greater than that necessary to encode the
188 amino acid protein. Primer extension analysis of
Salmonella RNA using oligonucleotide primers specific for
pagC sequence was performed to determine the approximate
start site of transcription and to determine whether
these nucleotides might be transcribed 5' or 3' to the

- 52 -

188 amino acid *pagC* gene product. Primer extension analysis with an oligonucleotide predicted to be complementary to nucleotides 550-565 of *pagC*, 150 nucleotides 5' to the predicted start codon, resulted in

5   an approximately 300 nucleotide primer extension product. Therefore a primer further upstream was constructed complementary to nucleotides 335-350 of *pagC* and used in a similar analysis. A primer extension product of 180 nucleotides was observed to be primer specific. This is

10   consistent with transcription starting at nucleotide 170 (Fig. 3). Upstream of the predicted transcriptional start, at nucleotides 153-160, a classic RNA polymerase binding site was observed with the sequence TATAAT at -12 nucleotides as well as the sequence TAATAT at -10

15   nucleotides. No complete matches were observed for the consensus RNA polymerase recognition site (TTGACA) 15-21 nucleotides upstream from the -10 region. AT -39 (126-131) nucleotides (TTGGAA), -38 (127-132) nucleotides (TTGTGG), and -25 (135-140) nucleotides (TTGATT) are

20   sequences that have matches with the most frequently conserved nucleotides of this sequence.

Based on the above results transcription was predicted to terminate near the translational stop codon of the 188 amino acid protein (nucleotide 1295, Fig. 3).

25   Indeed, a stem loop configuration was found at nucleotides 1309-1330 that may function as a transcription terminator. This was consistent with the lack of evidence of open reading frames downstream of the 188 amino acid protein and the lack of synthesis of other

30   transcription/translation using the cloned *pagC* DNA. This further suggests that the *pagC*::Tn*phoA* insertion inactivated the synthesis of only a single protein.

- 53 -

Similarity of pagC to Ail and Lom

A computer analysis of protein similarity using
the National Biomedical Research Foundation/Protein
Identification Resource, George et al., 1986, Nucleic
Acids Res. 14:11-15, hereby incorporated by reference,
protein sequence base was conducted to identify other
proteins that had similarity to pagC in an attempt to
find clues to the molecular function of this protein.
Remarkably, pagC was found to be similar to a
bacteriophage lambda protein, Lom, that has been
localized to the outer membrane in minicell analysis,
Court et al., 1983, Lambda II, Hendrix, R.W. et al. ed.
Cold Spring Harbor Laboratory (Cold Spring Harbor NY),
pp. 251-277, hereby incorporated by reference, and
demonstrated to be expressed by lambda lysogens of E.
coli, Barondess, et al., 1990, Nature 346:871-874, hereby
incorporated by reference. Recently, the deduced amino
acid sequence of the cloned ail gene product of Y.
enterocolitica was determined and found to also be
similar to Lom, Miller et al., 1990b, J. Bacteriol.
172:1062-1069. Therefore, a protein family sequence
alignment was performed using a computer algorithm that
establishes protein sequence families and consensus
sequences, Smith et al., 1990, Proc. Natl. Acad. Sci.
87:118-122, hereby incorporated by reference. The
formation of this family is indicated by the internal
data base values of similarity between these proteins :
pagC and Lom (107.8), pagC and Ail (104.7), and Ail and
Lom (89.8). These same proteins were searched against
314 control sequences in the data base and mean values
and ranges were 39.3 (7.3-52.9) pagC, 37.4 (7.3-52.9)
Ail, and 42.1 (7.0-61.9) Lom. The similarity values for
this protein family are all greater than 3.5 standard
deviations above the highest score obtained for
similarity to the 314 random sequences. No other

- 54 -

similarities or other family members were found in the
database.   Regions of similarity are located not only in
the leader peptide transmembrane domains but throughout
the protein.

5   _pag_ Mutant Strains Are Attenuated For Virulence
              Salmonella typhimurium strains of the invention
with a _pagC_ mutation were attenuated for virulence by
least 1,000-fold.
              In addition _pagC_, other _pag_ genes described herein
10   may be useful in the development of live _Salmonella_
vaccines.   Mutations in _phoP_-activated genes could be
used to construct attenuated, live _Salmonella_ vaccines.
In constructing multivalent _Salmonella_ vectored vaccines,
PhoP-activated promoters could increase immunogenecity by
15   targeting foreign protein expression to antigen
presenting macrophages.
     Identification of novel _phoP_-activated genes.
              To further analyze the role of _phoP_-activated
genes in bacterial virulence, a bank of strains with
20   active _phoA_ gene fusions was generated by Tn_phoA_
mutagenesis.   Strain CS019 was the parent strain for
Tn_phoA_ mutagenesis because it has wild-type bacterial
virulence and carries the _phoN2_ allele, which results in
minimal background phosphatase activity.   Strains with
25   active _phoA_ gene fusions were identified by blue colony
phenotype after growth in agar containing XP.   Such
strains were then screened for decreased fusion protein
activity on acquisition of the _phoP12_ allelle that
results in a PhoP-null phenotype.
30   Two thousand and sixty-four AP expressing strains
were isolated and colony purified from two hundred and
forty independent matings. Strains with AP activity were
isolated at a frequency of 0.8% from the total pool of
kanamycin resistant (Tn_phoA_ containing) bacteria.   A
35   total of fifty-four candidate _pag_::Tn_phoA_ insertions were

- 55 -

isolated from the AP expressing strain bank, and forty-
nine of these were determined to have greater than six-
fold reduction in AP activity in the absence of
functional *phoP/phoQ*. Therefore, approximately 2% of the
5   colonies expressing AP were identified as *pag-phoA* gene
fusions.

Identification of thirteen unique *pag* loci.

        Three methods were used to determine whether the
forty-nine Tn*phoA* insertions defined unique *pag* loci.
10   First, physical maps of the *EcoRI* and *HindIII* restriction
endonuclease sites 5' to the Tn*phoA* insertions were
defined. Second, linkage analysis to transposon
insertions highly linked to known *pag* loci was performed.
Third, strains determined to be unique by the above
15   methods were screened for linkage to a bank of strains
with transposon insertions of known chromosomal location.

        Blot hybridization analysis demonstrated that
thirteen of the forty-nine strains had unique restriction
endonuclease sites 5' to the Tn*phoA* insertion. The
20   numbers of strains with similar physical maps 5' to the
Tn*phoA* insertion ranged from 1-7. One of the thirteen
physical maps was similar to that expected for an
insertion in *pagC* and was noted in seven of the strains
isolated as containing candidate *pag*::Tn*phoA* insertions.
25   Analysis of these seven strains indicated that only three
of these were *pagC*::Tn*phoA* insertions, since blot
hybridization analysis with a fragment of *pagC* as a probe
and linkage analysis to transposon insertions highly
linked to *pagC* indicated that four of these insertions
30   were not in *pagC*. Another of the *pag*::*phoA* fusions,
denoted *pagP*, had the physical 5' restriction-
endonuclease map that would be expected for *phoN*.
However, this insertion was determined not to be within
phoN by linkage analysis and blot hybridization. A
35   transductional cross was performed between wild type

- 56 -

bacteria and strain CS1247 containing *pagP*::Tn*phoA* and
zxx::6215Tn*10d-cam*. These transductants were selected on
kanamycin, insuring the inheritance of the *pagP*::Tn*phoA*
which encodes kanamycin resistance. These colonies were
5   then screened for choramphenicol resistance which would
indicate linkage of zxx:6215Tn*10d-cam* to *pagP*. No
linkage was found indicating that *pagP* was not linked to
phoN. Blot hybridization using a portion of *phoN* as a
probe was also performed on CS1247 and indicated that
10   this strain contained a wild type *phoN* locus. Thirteen
pag loci were defined and designated *pagD-P*.

    To further define the PhoP regulation of the 13
pag::Tn*phoA* fusion proteins, AP activity was assayed in
strains isogenic except for the *phoP* locus. AP activity
15   was assayed during bacterial growth in rich medium in
logarithmic and stationary growth phase (Table 13). The
dependence of an intact *phoP* locus for full expression
remained constant for the different stages of growth;
however, the relative amount of AP expression increased
20   as growth was limited. The difference in expression of
pag gene fusions varied from six to forty-eight fold when
isogenic strains with a wild type and null *phoP* locus
were compared.

    Of the five previously identified *pag* loci, only
25   *phoN*, *pagC*, and *pagA* have known chromosomal locations.
Linkage analysis of the 13 newly identified *pag* loci was
performed using strains containing transposon insertions
linked to *pagC* (AK3233 and AK3140), and to *pagA* (AK3255).
Three *pag*::Tn*phoA* insertions were found to be linked to
30   AK3140 which is in a region near *pagC* at 24-25 minutes on
the chromosome. These were designated *pagD*, *pagE*, and
pagF. *PagD*::Tn*phoA* was similarly linked to the
transposon insertion of AK3233 (83%) and AK3140 (33%) as
was previously reported for *pagC*. The Tn*phoA* insertion
35   of this strain has been further defined and is

- 57 -

divergently transcribed from *pagC*. *pagE* and *pagF*
exhibited different linkage to the insertions of AK3233
and AK3140 than *pagC* and *pagD* suggesting a significantly
different chromosomal location. The *pagE*::Tn*phoA*
5   insertion is 39% linked to the transposon insertion of
AK3233 and 99.1% linked to that of AK3140, while
pagF::Tn*phoA* is 31% linked to the insertion of AK3140 but
not to that of AK3233. These different linkages in
addition to the physical maps of the restriction
10   endonuclease sites 5' to the Tn*phoA* insertion indicated
that these were new *pag* loci. Therefore, three new *pag*
loci were found in the region of 25 minutes, one of which
is highly linked to the previously defined *pagC*.
         Linkage analysis was then performed using a group
15   of defined random Tn*10*Δ*16*Δ*17* insertions on the ten
strains with Tn*phoA* insertions of no known location. Of
these ten *pag*::Tn*phoA* alleles only two demonstrated
linkage to the bank of Tn*10*Δ*16*Δ*17* insertions. The
pagG::Tn*phoA* insertion was demonstrated to have 97%
20   linkage to the transposon insertion of AK3258 located at
approximately 30 minutes. The *pag*::Tn*phoA* insertion,
designated *pagH*, exhibited 23% linkage to the insertion
of AK3091. The linkage to the transposon insertion of
AK3091 was similar to linkage previously demonstrated for
25   *prgE* (26%). Therefore, this chromosomal region contains
both PhoP-activated and repressed genes. This TnΔ*16*Δ*17*
insertion was analyzed using pulse field gradient
electrophoresis of chomosomal DNA from AK3091 digested
with the restriction endonuclease *XbaI* and *BlnI*. These
30   data indicate that the transposon insertion of AK3091 was
located in the region of 20-25 minutes and that *pagH* and
prgE are located in this region of the chromosome.

- 58 -

Strains with pag:TnphoA insertions have wild type
sensitivity to the rabbit NP-1 defensin

        S. Typhimurium strains with null mutations in the
phoP operon have increased sensitivity to a variety of
5   cationic antimicrobial peptides including defensins,
magainins, and protamine. The defensins are a family of
mammalian peptides present in the granules of
neutrophils, lung macrophages, and intestinal Paneth
cells. Resistance to these peptides may contribute to
10   bacterial virulence and the ability to colonize mucosal
surfaces. Strains with pag::TnphoA insertions were
tested for sensitivity to the highly active rabbit
defensin NP-1. None of the strains with single
pag::TnphoA insertions demonstrated increased sensitivity
15   to NP-1 defensin (see Fig. 6). Thus despite the
demonstrated sensitivity of PhoP-null mutants to rabbit
defensin NP-1, no defined mutations in pag loci were
associated with sensitivity to defensins.

Four strains with pag::TnphoA insertions demonstrate
20   marked attenuation for mouse virulence

        To further define whether these new pag loci
contributed to mouse virulence, the 13 strains with pag
transposon insertions were screened in vivo. Mice were
injected intraperitoneally with approximately 100
25   organisms. Four strains with transposon insertions in
pagD, pagJ, pagK, and pagM demonstrated attenuated
virulence. Mice injected with these strains all survived
and showed no signs of systemic infections, such as
hepatosplenomegaly and scruffiness (piloerection due to
30   fever). These four strains were subjected to further
virulence testing by intraperitoneal injection of
multiple doses of organisms in a total of ten mice on two
separate occasions. The mean LD_{50} was determined from
these subsequent injections and is listed in Table 14.
35   One of these strains, containing the pagD::TnphoA
insertion, has a LD_{50} 10,000 fold greater than wild-type

- 59 -

S. typhimurium. The other three strains were also markedly attenuated for mouse virulence with LD_{50} values greater than 1000-10,000 times that of wild type organisms. These data indicated that the PhoP-regulated

5 loci, pagD, pagJ, pagK, and pagM, when mutated, result in attenuation of bacterial virulence.

<u>pag::TnphoA strains attenuated for mouse virulence have reduced survival within macrophages.</u>

Since PhoP mutant *Salmonella* are deficient in

10 survival within macrophages, strains containing mutations in *pag* genes that had attenuated mouse virulence were tested for reduced viability within macrophages. As shown Table 14, all strains with *pag* mutations demonstrated significantly reduced survival within

15 macrophages. Decreased intracellular survival of *pag* mutants was not observed until a time when *pag* are predicted to be maximally expressed.

Four strains with mutations in the *pagC, pagD, pagJ, pagK* and *pagM* loci were found to be attenuated for

20 mouse virulence and survival within macrophages. Strains with mutations in these five *pag* all had varying degrees of virulence attenuation. Strains with a mutation in *pagJ* had a virulence defect comparable to that observed for *pagC* mutants (greater than 1000 x the LD_{50} of wild

25 type organisms). The *pagD::TnphoA* insertion resulted in the greatest attenuation of virulence, comparable to that of a PhoP null mutation (greater than 10,000 x the LD_{50} of wild type organisms). *pagK* and *pagM* mutants had virulence attenuation that was intermediate between the

30 *pagJ* and *pagD* mutants. The cumulative effect of deletion of *pagC, pagD, pagJ, pagK,* and *pagM*, if additive and similar to the attenuation observed with TnphoA insertions, may be much greater than that observed by deletion of *phoP* alone. The observation that many of

35 these genes are somewhat expressed in stationary phase in the absence of PhoP suggests that functional Pag proteins

- 60 -

could be produced *in vivo* in the absence of PhoP. One
virulence gene *pagM* is significantly expressed in the
absence of PhoP, though it may still require PhoP/PhoQ
for induction within macrophage phagosomes. This data
5   suggests that deletion of *pag* gene products could lead to
greater virulence attenuation than deletion of the
regulatory proteins.

Salmonella envelope proteins as virulence factors:
Defensin senstivity

10   Based on the methods used to identify *pag* loci,
i.e., translational gene fusions to *phoA*, and the
observation that the pagC gene fusions produce AP, it has
now been discovered that many *pag* encode bacterial
envelope proteins. No strains have been found with
15   single *pag* mutations that confer sensitivity to defensins
or other cationic peptides. The data suggest that an
alteration of the bacterial envelope as a result of the
change in synthesis of the entire aggregate of envelope
proteins mediated by PhoP/PhoQ may be important to *S.*
20   *typhimurium* virulence.

          Defensins are small amphipathic cationic peptides
of approximately 30-35 amino acids in length whose anti-
microbial action involves penetration and disruption of
membranes, possibly by forming selective anionic
25   channels. Though defensins are largely found in
neutrophils and Paneth cells these or other related
molecules likely contribute to non-oxidative killing of
phagocytosed bacteria by macrophages. Though it remains
possible that a single unidentified *pag* encodes a protein
30   responsible for defensin resistance, it seems more likely
that a cumulative effect of expression of several *pag*
encoded envelope proteins could result in resistance to
defensins. An aggregate change in a large number of
bacterial envelope proteins could alter the membrane
35   charge, electrical potential, or lipid content such that
defensin interaction with bacterial membranes could be

- 61 -

changed.

<u>Identification of transcriptional units linked to <i>pagC</i>.</u>

     To identify genes upstream of <i>pagC, E. coli</i>
carrying plasmid pWPL17 containing 2.8kb of DNA 5' to
5   <i>pagC</i> (Table 15 and Fig. 7) was mutagenized with the
transposons Mu<i>dJ</i> and Tn<i>phoA</i>, and strains with AP or β-
galactosidase activity were identified on chromogenic
substrates.   In addition, as part of an effort to
identify additional PhoP-activated genes, random
10   mutagenesis of the <i>Salmonella</i> chromosome with Tn<i>phoA</i> was
performed, and strains with AP activity were screened for
Tn<i>phoA</i> insertions linked to the Tn<i>10Δ16Δ17</i> of strain
AK3233, which is 75% linked to <i>pagC</i>.   Several strains
that contained plasmids with active Mu<i>dJ</i> or Tn<i>phoA</i>
15   generated gene fusions were identified.   In addition, two
strains were identified that contained active chromosomal
Tn<i>phoA</i> insertions closely linked to <i>pagC</i>.   Physical maps
of the restriction endonuclease sites surrounding the
transposon insertions in strains with active plasmid or
20   chromosomal <i>lacZ</i> and <i>phoA</i> gene fusions were performed to
determine the relationship of the transposon insertions
to <i>pagC</i>.   This analysis revealed that several regions of
the DNA were transcribed oppositely to <i>pagC</i> (Fig. 7).
Several Tn<i>phoA</i> insertions that resulted in active <i>phoA</i>
25   gene fusions were identified.   These data indicated that
<i>pagC</i>-linked genes encoded membrane or secreted proteins.

<u>Genes linked to <i>pagC</i> encode four novel proteins.</u>

     To further analyze the genes defined by transposon
insertions, the DNA sequence of this region was
30   determined (Fig. 8).   DNA containing this region was
cloned; 4 kb of DNA between the <i>Hpa</i>I site 737bp upstream
of the start codon of <i>pagC</i> to a <i>Cla</i>I site far upstream
was sequenced.   The DNA sequence of the fusion junctions
of all Tn<i>phoA</i> and Mu<i>dJ</i> gene fusions was also determined.
35   Based on these data, the correct reading frame of each

- 62 -

gene was determined. The DNA sequence data revealed four
ORFs predicted to be transcribed and translated based on
the data derived from the TnphoA and MudJ insertions.
All ORFs revealed typical ribosome binding sites 6 to 11

5   bases from the predicted start of translation. The
translation of the ORF immediately upstream and
oppositely transcribed to pagC, pagD, indicates that a
short envelope protein of 87 amino acids (unprocessed) is
encoded. It is followed by a second ORF (envE) which

10   encodes an envelope protein of 178 amino acids
(unprocessed). This ORF is followed by a structure that
could function as a Rho-independent transcriptional
terminator (see Fig. 8). The third ORF, msgA (macrophage
survival gene), encodes a small protein similar in size

15   to that of the first gene product (79 amino acids) and is
also followed by a structure that could function as a
Rho-independent transcriptional terminator (see Fig. 8).
The DNA sequence predicts that this protein is composed
of several charged residues with a large number of

20   negatively charged amino acids residing at the carboxy
terminus. The predicted protein product does not contain
a structure resembling a signal sequence at its amino
terminus nor any hydrophobic stretches; therefore, the
third ORF is unlikely to encode an envelope protein. The

25   final ORF (envF) encodes an envelope protein of 278 amino
acids (unprocessed). A computer search of known protein
motifs revealed that EnvF contains a consensus
prokaryotic membrane lipid attachment site and,
therefore, is likely to be a lipoprotein (see Fig. 8 for

30   consensus site location).

The predicted proteins produced by pagD, envE, and
envF contain a typical bacterial signal sequence
structure. In addition, hydrophobic profiles confirmed
the hydrophobic nature of the amino-termini of these

35   proteins. The EnvE and EnvF proteins also contain

- 63 -

hydrophobic stretches that could function as membrane
spanning domains. The G+C content of the genes in this
region are: *pagC*, 43.4%; *pagD*, 42.1%; *envE*, 45.9%; *msgA*,
46.8%; and *envF*, 40.5%, which is considerably lower than
5   the average G+C content of *S. typhimurium* (52%). A
complete search of the database with the predicted
protein sequences of these four ORFs showed no
significant similarities. Strains containing three
distinct Tn*phoA* insertions and one Mu*dJ* insertion, each
10   located in one of the four genes, were chosen for further
characterization.

A gene pagD, oppositely transcribed to pagC, is
positively regulated by PhoP/PhoQ

Representative strains with transposon insertions
15   were examined to evaluate whether genes transcribed
oppositely to *pagC* were increased in synthesis in the
presence of PhoP. To accurately determine if these genes
were PhoP regulated, it was necessary to recombine
plasmid insertions onto the *Salmonella* chromosome. Upon
20   replacement of the wildtype gene with the gene containing
the transposon insertion, P22HT*int* lysates made on these
strains were transduced into a PhoP deleted (PhoP⁻)
strain and AP or β-galactosidase levels were monitored.
One of these transposon generated gene fusions
25   demonstrated a significant increase in activity between
PhoP⁻ and WT backgrounds, while the other insertions
showed no PhoP regulation (Table 16). This *pagD* loci is
adjacent to and divergently transcribed from *pagC*.

The representative transposon insertion in *envF*
30   was unable to be recombined onto the chromosome, likely
due to an insufficient amount of homologous DNA
downstream of the transposon. In order to examine the
possibility of PhoP regulation of the *envF* gene, a region
upstream of this gene through and including the *phoA* gene
35   of the Tn*phoA* transposon was cloned as a 3-kb *Pvu*I
(blunt-ended)-*Xho*I fragment into the *Eco*RV-*Sal*I sites of

- 64 -

the suicide vector pGP704. This clone was mated into
Salmonella strain CS019, and ampicillin-resistant
recombinants were selected (creating a strain designated
envF::pGPP2). A *phoP*105::Tn10d-Tet mutation was

5   transduced into this strain to create an isogenic pair
differing only in the ability to produce a functional
PhoP protein. As shown in Table 16, the introduction of
the *phoP*105::Tn10d-Tet had no effect on the AP levels of
these two strains, demonstrating that *envF* is not a PhoP-

10   activated gene.

Transposon insertions in *pagC*-linked genes attenuate
virulence and cause reduced survival within macrophages
        Since transposon insertions in *pagC* significantly
increase the LD$_{50}$ of *S. typhimurium* in BALB/c mice,

15   strains containing transposon insertions linked to *pagC*
were evaulated for attenuation of mouse virulence. As
shown in Fig. 7, while the transposon insertion in *envE*
had no affect on strain virulence, a Tn*phoA* insertion in
pagD and the Mu*dJ* insertion 1.8 kb downstream in *msgA*

20   attenuate *S. typhimurium* virulence by greater than 300
fold as compared to wild-type organisms (LD$_{50}$<20
organisms). These data suggested that these two loci are
essential to virulence.
        To examine the survival capabilities of those

25   strains having a virulence defect, *S. typhimurium*
containing insertions in either *pagD* or *msgA* were used to
infect bone marrow-derived macrophages. The results,
shown in Table 15, demonstrate a macrophage survival
defect for these two strains. The survival defect is

30   greater for the *pagD* insertion (MSI=0.002) compared with
the *msgA* insertion (MSI=0.01), and both defects are equal
to or greater than that of the PhoP⁻strain (MSI=0.01).
        Transposon insertions in this gene could not be
recombined onto the chromosome. Thus, it was necessary

35   to demonstrate that the virulence and macrophage survival
defects of *msgA* was not due to a polar effect of the Mu*dJ*

- 65 -

insertion on *envF* transcription. Therefore, pGPP2 was
recombined into the *msgA*::Mu*dJ* strain and AP activity of
this strain was compared to that of CSO19 containing the
recombinant pGPP2. This data (shown in Table 16)
5   demonstrates that the transcription of the *envF* gene is
unaffected by the *msgA*::Mu*dJ* insertion and is transcribed
from its own promoter. However, it is possible that
under different environmental conditions, other promoters
may be activated that could place *msgA* and *envF* on the
10   same transcript.

Determination of the *msgA* and *pagD* transcriptional start
sites

        The 5' regions of these genes were examined to
define the transcriptional start sites of *msgA* and *pagD*.
15   Oliogonucleotides complimentary to the 5' end of each ORF
or upstream region were used in a primer extension
analysis. The results of this analysis revealed that the
pagD transcript begins 39 bases upstream of the
translational start. The predicted -10 (TTCCAT) and -35
20   (TTGAAT) regions were found to be similar to the known
consensus sequences for *E. coli* promoters. The *pagD*
transcript was detected only in PhoP^c *Salmonella* RNA and
not in RNA from PhoP$^-$ *Salmonella*. The *msgA*
transcriptional start was found to begin 58 bases
25   upstream of the translational start and contain predicted
-10 (CAAAAC) and -35 (TTACGT) sequences. These regions
do not conform well to consensus -10 and -35 sequences;
however, the cDNA from this transcript was easily
detected using various primers in primer extensions of
30   both PhoP^c and PhoP$^-$ RNA and appears to produce an
abundant RNA.

Distribution of *pagD* and *msgA* genes in the
Enterobacteriaceae and in two G+C content organisms

        The G+C content of the *pagC* chromosomal region is
35   much lower than the average G+C content of *Salmonella*.
The gene encoding the PhoP-regulated acid phosphatase of

- 66 -

S. typhimurium (phoN) also has a low G+C content (39%),
and DNA homologous to *phoN* was found only in two low G+C
organisms of several genera tested. The DNAs of several
members of the Enterobacteriaceae and two low G+C
5   organisms were examined for similarity to *pagD* and *msgA*
by blot hybridization. PCR fragments highly specific to
each ORF were labeled and used as probes. This analysis
demonstrated hybridization at high stringency to all
Salmonella species examined as well as *Shigella sonnei,*
10   *Shigella flexneri, Klebsiella pneumoniae* and *Citrobacter
freundii.* No hybridization was seen to the low G+C
organisms *Morganella morganii* or *Providencia stuartii.*
Identical hybridization patterns were seen with probes
specific for both genes indicating that these genes are
15   also linked in organisms other than *Salmonella.*

A virulence gene cluster required for *Salmonells
typhimurium* survival within macrophage macrophages

Four genes upstream and oppositely transcribed to
the *pagC* gene of *Salmonella typhimurium* have now been
20   identified. Three genes (*pagD, envE* and *envF*) are
predicted to be envelope proteins based on the isolation
of active Tn*phoA* insertions in these loci and the
presence of a typical signal sequence at the amino-
terminus of each protein. None of the four proteins
25   possess significant homology to any protein in the
database.

Only the gene immediately upstream of *pagC* and
oppositely transcribed (*pagD*) was determined to be PhoP
regulated. Transposon insertions in this gene greatly
30   attenuate virulence and the ability of the organism to
survive within murine macrophages. The transcription of
several *pag* (including *pagC*) has been shown to be induced
when *Salmonella* are within macrophage phagosome. In
addition, analysis of proteins produced by *Salmonella*
35   after infection of macrophage-derived cell lines indicate
that *pag* products are induced and that *pagC* may be among

- 67 -

the most abundant gene products induced upon macrophage infection. Since *pagD* is required for macrophage survival, it is likely that the transcription of this gene also will be induced within macrophage phagosomes.
5 The *pagD* protein is small (87 amino acids, unprocessed) and has no strong hydrophobic domains; therefore, it is likely that it is a periplasmic or secreted protein.

Transposon insertions in the *msgA* gene were found to have an effect on mouse virulence and macrophage
10 survival. It is likely that this gene may also be induced within acidified macrophage phagosomes as are other genes necessary for macrophage survival. If this gene is induced by the macrophage environment, its expression (as well as other genes necessary for
15 macrophage survival) may be controlled by a regulatory system separate from the PhoP/PhoQ system.

These *pagC*-linked genes do not appear to form an operon. Because none of the genes downstream of *pagD* are PhoP regulated, they appear not be transcribed from the
20 *pagD* promoter. The presence of a potential transcriptional terminator at the end of the *envE* gene makes it unlikely that *msgA* is co-transcribed with *envE*. The data suggest that the *msgA*::Mu*dJ* insertion is not polar on *envF*, which suggests that *envF* has its own
25 promoter. Additionally, a potential transcriptional terminator following *msgA* as well as a 493 bp intergenic region makes it unlikely that these genes are co-transcribed. Primer extension analysis of these genes confirms that all four genes are transcribed from their
30 own promoter.

The other two genes identified in this region, *envE* and *envF*, appear to produce membrane proteins that contain characteristic membrane spanning regions. The *envF* gene product is likely to be a lipoprotein based on
35 the presence of a consensus lipid attachment site and is

- 68 -

likely to play a role in *Salmonella* virulence.

The low G+C content of the genes in the *pagC*
region suggests that they may have been acquired by
horizontal transmission. Southern blot analysis of low

5   G+C organisms probed with the *msgA* or *pagD* genes showed
no homology, but this does not eliminate the possibility
that they were acquired from another low G+C content
organism. The possibility also exists that these genes
reside on a mobile genetic element acquired from another

10 source. The *msgA* and *pagD* probes hybridized in identical
patterns to some members of the Enterobacteriaceae other
than *Salmonella*. However, the *pagC* gene has been shown
to be unique to *Salmonella* species. This may indicate
that the products of the genes upstream of *pagC* do not

15 form a complex with PagC or that their functions do not
require PagC interaction. Alternatively, because
proteins that have homology to PagC exist in other
Enterobacteriaceae (in the absence of any DNA homology),
a PagC homolog may be linked to *msgA* and *pagD* in other

20 species which was not detected by the DNA hybridization
experiments.

pagC/pagD promoter region: expression of heterologous
proteins

pagC and *pagD* are divergently transcribed and are

25 both PhoP activated. Other divergently transcribed,
regulated genes are known in the art (Beck et al., 1988,
Microbiol. Rev. 52:318-326), e.g., the *Klebsiella
pneumoniae pulA-malX* region (Chapon et al., 1985, J.
Bacteriol. 164:639-645). Transcription of most of such

30 genes require accessory proteins, such as CAP, in
addition to the regulator to activate transcription.
These two genes are divergently transcribed, and their
promoters are arranged back-to-back. A region of 134 bp
exists between transcriptional start sites of these

35 genes, which is similar to the intergenic region between
pagC and *pagD*. The *pulA-malK* promoter region is

- 69 -

predicted to contain two MalT (the regulatory protein of
this system) binding sites, one for each gene. Other
MalT-activated genes require the CAP protein for
expression, but the *pulA* and *malX* genes do not, possibly
5   because of the high local concentration of the MalT
regulator. Since the region between the transcriptional
start sites of *pagC* and *pagD* (the predicted -35
sequences) is only 137bp (nucleotides 562 to 776 of SEQ
ID NO:15), it is likely that only PhoP binding sites
10   exist in the intergenic region, and that binding of one
or more phosphorylated PhoP molecules positively
regulates both genes. This *pagC/pagD* intergenic region
which contains the divergent promoters can be used to
construct vectors to express two heterologous proteins,
15   one in each direction.

prg genes

As discussed above, *phoP/phoQ* constitutive
mutations (phenotype PhoP^c) increase the expression of
pag and repress the synthesis of approximately 20
20   proteins encoded by *phoP*-repressed genes (*prg*). PhoP^c
bacteria are attenutated for mouse virulence suggesting
that *prg* are virulence genes.

By use of the transposon, Tn*phoA*, five unlinked
prg loci were identified. In general, media conditions
25   (starvation) that activate *pag* expression repress *prg*
expression. One *prg* locus, *prgH,* was demonstrated to
contribute to mouse virulence by both the oral and the
intraperitoneal route. Both PrgH as well as PhoP^c mutant
S. typhimurium were found to be defective in induction
30   of endocytosis by epithelial cells. Identification and
mutation of such virulence genes will be useful in
vaccine development.

- 70 -

Nucleotide sequence of the prg H, prgI, prgJ, and prgK
genes

　　　　SEQ ID NO:10 represents the nucleotide sequence of
a 5100-bp HindIII fragment that contains the

5　hyperinvasive hil locus. Four ORFS encoding four prg
genes are located within this DNA (see Fig. 9). The ATG
start codon is underlined; the asteriks indicate the
positions of the prgH, prgI, prgJ, and prgK stop codons.
These prg loci are required for bacterial invasion of

10　epithelial cells, full mouse virulence, and
transepithelial neutrophil migration. A bacteria
attenuated by a mutation in one or more of these loci can
be used to vaccinate individuals against infection by the
wild type pathogen.

15　Strains, materials and methods

　　　　All bacterial strains used in the characterization
of prg genes are listed in Table 5.

- 71 -

Table 5

5	Strain genotype or description	Relevant Reference or source
	S. typhimurium	
10	14028s derivatives	
	14028s Wild type	ATCC
	CS002 *phoP12*	This work
	CS003 Δ*phoP* Δ*purB*	This work
	CS012 *pagA1::Mu dJ*	This work
15	CS013 *pagB1::Mu dJ*	This work
	CS119 *pagC1::TnphoA phoN2 zxx::6251 Tn10d-Cm*	This work
	CS015 *phoP-102 ::Tn10 d-Cm*	This work
	CS019 *phoN2 zxx::6251Tn10d-Cm*	This work
	CS022 *pho-24*	This work
20	CS023 *pho-24 phoN2 zxx::6251Tn10d-Cm*	This work
	CS030 *phoN2 zxx::6251Tn10d-Cm phoP12*	This work .
	AD154 *phoP12 purB1744::Tn10*	Gift of E. Eisenstadt
	CS031 *pho-24 purB1744::Tn10*	This work
25	IB001 *phoN2 zxx::6251Tn10d-Cm* Δ*phoP* Δ*purB*	This work
	IB002 CS030 with *prgA1::TnphoA*	This work
	IB003 IB002 with *pho-24 purB1744::Tn10*	This work
	IB004 IB002 with *phoP12 purB1744::Tn10*	This work
	IB005 CS019 with *prgA1::TnphoA*	This work
30	IB006 CS015 with *prgA1::TnphoA*	This work
	IB007 CS030 with *prgB1 ::TnphoA*	This work
	IB008 IB007 with *pho-24 purB1744::Tn10*	This work
	IB009 IB007 with *phoP12 purB1744::Tn10*	This work
	IB010 CS019 with *prgB1 ::TnphoA*	This work
35	IB011 CS015 with *prgB1 ::TnphoA*	This work
	IB012 CS030 with *prgB2::TnphoA*	This work
	IB013 IB012 with *pho-24 purB1744::Tn10*	This work
	IB014 IB012 with *phoP12 purB1744::Tn10*	This work
	IB015 CS019 with *prgB2::TnphoA*	This work
40	IB016 CS015 with *prgB2::TnphoA*	This work
	IB017 CS030 with *prgC1::TnphoA*	This work
	IB018 IB017 with *pho-24 purB1744::Tn10*	This work
	IB019 IB017 with *phoP12 purB1744::Tn10*	This work
	IB020 CS019 with *prgC1::TnphoA*	This work
45	IB021 CS015 with *prgC1::TnphoA*	This work
	IB022 CS030 with *prgE1::TnphoA*	This work
	IB023 IB022 with *pho-24 purB1744::Tn10*	This work
	IB024 IB022 with *phoP12 purB1744::Tn10*	This work
	IB025 CS019 with *prgE1::TnphoA*	This work
50	IB026 CS015 with *prgE1::TnphoA*	This work
	IB027 CS030 with *prgE2::TnphoA*	This work
	IB028 IB027 with *pho-24 purB1744::Tn10*	This work

- 72 -

	IB029	IB027 with *phoP12 purB1744*::Tn*10*	This work
	IB030	CS019 with *prgE2*::TnphoA	This work
	IB031	CS015 with *prgE2*::TnphoA	This work
	IB032	CS030 with *prgE3*::TnphoA	This work
5	IB033	IB032 with *pho-24 purB1744*::Tn*10*	This work
	IB034	IB032 with *phoP12 purB1744*::Tn*10*	This work
	IB035	CS019 with *prgE3*::TnphoA	This work
	IB036	CS015 with *prgE3*::TnphoA	This work
	IB037	IB001 with *prgH1*::TnphoA	This work
10	IB038	IB037 with *pho-24 purB1744*::Tn*10*	This work
	IB039	IB037 with *phoP12 purB1744*::Tn*10*	This work
	IB040	CS019 with *prgH1*::TnphoA	This work
	IB041	CS015 with *prgH1*::TnphoA	This work
	IB042	Tn5B50-380 in IB040	This work
15	IB043	pWKSH5 in IB040	This work
	IB044	pWKSH5 in CS022	This work
	CS032	*oxiA1049*::Mu d1-8 *supD10*	This work
	CS033	*oxiC1048*::Mu d1-8 *supD10*	This work
	CS034	*oxiE4*:: Mu d1 Δ*nadA100*	This work

20   Other *S. typhimurium* derivatives

	AK3011-AK3314		Collection
	of randomly spaced Tn*10*Δ16Δ17 insertions		(19)
	TT520	*srl-202*::Tn*10*	(41)
	TT2979	*srl-211*::Tn*5*	(41)
25	TN3061	*zcf-845*::Tn*10* *dcp-1* *zhg-1635*::Tn*10*dCm	(41)
	SH7782	*ompD*::Tn*5*	(41)
	ₓ4115	*invA*::*cat*	(13)
	EE517	Δ*hil-517* (Tn5B50-380)	Gift of C.
	Lee		
30	JF897	*oxiA1049*::Mu d1-8 *supD10*	(2)
	JF896	*oxiC1048*::Mu d1-8 *supD10*	(2)
	JF739	*oxiE4*::Mu d1 Δ*nadA100*	(2)

S. enteritidis

| | CDC5 | clinical wild-type isolate | (45) |
| 35 | SM7 | Str^r *smb* | (45) |

E. coli

	SM10(pRT291)		contains
	plasmid pRT291 (TnphoA) derived from		(49)
		pRK290 selecting for Tc^r and Km^r.	
40	MM294(pPH1JI)	contains Gm^r plasmid pPH1JI, which is incompatible	
	(49)		
		with pRK290	
	VV42(pWKSH5)		contains
	plasmid pWKSH5, a derivative of pSC101		
45		(51) that contains a 5.1 kb *HindIII* fragment of *hil* DNA	
	including prgH		V.Bajaj and
			C.Lee

- 73 -

(19)   Kukral et al., Journal of Bacteriology, 169:1787-
        1793, 1987

(41)   Sanderson et al., Microbiological Reviews, 52:485-
        532, 1988

5 (13)   Galan et al., Infection and Immunity, 59:3116-
        3121, 1990

(2)   Aliabadi et al., Journal of Bacteriology, 165:780-
        786, 1986

(45)   Stone et al., Journal of Bacteriology, 174:3945-
10   3952, 1992

        Bacteria were grown as follows: Luria-Bertani (LB)
broth was used as rich medium. Antibiotics were used in
the following concentrations in growth media or agar:
ampicillin 100 μg/ml (Ap), chloramphenicol 25 μg/ml (Cm),
15 gentamicin 30 μg/ml (Gm), kanamycin 45 μg/ml (Km), and
tetracycline 25 μg/ml (Tc). The chromogenic substrate
5-bromo-4-chloro-3-indolyl-phosphate (p-toluidine salt)
(XP) was used to detect phosphatase activity on agar at a
final concentration of 40 μg/ml. p-nitrophenyl
20 phosphate (p-NPP) was used as a substrate for
quantitative measurement of AP activity. Media was
buffered to various pH ranges with 1 M sodium citrate. E
media (Vogel-Bonner minimal) was prepared as described by
Davis et al., 1980, Advanced Bacterial Genetics: A
25 Manual for Genetic Engineering. Cold Spring Harbor
Laboratory, Cold Spring Harbor, N. Y. Nitrogen-, carbon-
, and phosphate free medium (N$^-$C$^-$P$^-$) was prepared as
described by Kier et. al., 1977, J. Bacteriol. 130:399,
herein incorporated by reference.
30 This starvation medium was supplemented with 0.04%
(wt/vol) glucose as the carbon source, 10 mM NH$_4$Cl as the
nitrogen source, and 1 mM NaH$_2$PO$_4$.H$_2$O as the phosphate
source. The carbon concentration is one log less than
described by Kier et al., $supra$.
35   AP activity of strains isogenic except for

- 74 -

mutations in the *phoP* locus was measured in cultures
grown from a single colony inoculum under various oxygen
tensions with or without shaking at 37°C. Anaerobic
cultures were grown in an anaerobic chamber (Coy

5   Laboratories Products, Inc.) with a gas mixture of 80%
N_2, 10% O_2, and 10% CO_2 at 37°C.   For acid regulation,
aliquots of mid-logarithmic cultures were removed to
measure initial pH and AP activity.   1M sodium citrate
(pH >6.0) or 1M citric acid (pH 4.7) were added to

10   equivalent amounts of culture to a final concentration of
50 mM citrate.   Cultures were grown aerobically for two
hours at 37°C and then pH and AP measurements were taken.
AP activity was measured as described previously
(Michaelis et al., 1983, J. Bacteriol. 154:366-374,

15   herein incorporated by reference).   AP units were
calculated by the following formula: units = {OD_{420}/[time
(minutes) x volume x OD_{600}]} x 1000 as defined by Miller
for β-galactosidase (Miller et al., 1972, Experiments in
molecular genetics, p. 352-355. Cold Spring Harbor

20   Laboratory, Cold Spring Harbor, NY.).

        Standard bacterial genetic techniques were used to
study *prg* loci.   Bacteriophage P22HT*int*-mediated
transduction was performed as according to methods known
in the art.   Tn*phoA* mutagenesis was performed using a

25   broad host range plasmid (pRT291) to deliver Tn*phoA*
(Taylor et al., 1989, J. Bacteriol. 171:1870, herein
incorporated by reference).   Transpositions of Tn*phoA*
into *Salmonella* DNA were identified by use of the
incompatibility plasmid pPH1JI (Taylor et al., *supra*).

30   Screening for *phoP*-repressed genes was performed using
CS031, the donor strain of the *pho-24* allele.   CS031 was
constructed by a P22 bacteriophage transductional cross
between strains AD154 and CS022 which contains the
purB::Tn*10* allele and the *pho-24* allele, respectively.

35   The linkage of *pho-24* and *purB*::Tn*10* was 70%, similar to

the linkage of *purB* to other *phoP* alleles. Therefore, when P22 bacteriophage transductional crosses were performed between CS031 and the strains containing active gene fusions to *phoA*, strains could be screened for loss

5   of fusion protein activity on acquisition of tetracycline resistance. Initial screening involved detection of loss of AP activity in approximately 70% of colonies that acquired tetracycline resistance, as they were presumed to contain the *pho-24* allele. In addition, controls

10   were performed using strain AD154 that contains the same *purB*::Tn*10* allele linked to a *phoP* null allele, *phoP12*. Plasmid DNA was transformed into *S. typhimurium* strain LB5010 by the calcium chloride and heat shock procedure (Maclachlan et al., 1985, J. Bacteriol. 161:442).

15   Isolation of strains with Tn*phoA* insertions in *phoP*-repressed genes

     Constitutive mutations in the *phoP* locus (phenotype PhoP^c) that result in increased expression of *pag* in an unregulated fashion also markedly attenuate *S.*

20   *typhimurium* virulence and survival within macrophages. The virulence defect of PhoP^c strains can be explained by their decreased expression of approximately 20 polypeptides encoded by *phoP*-repressed genes (*prg*).

     A PhoP^−PhoN^− strain (IB001) was constructed by a

25   P22 transductional cross between CS019 and CS003. IB001 was then mutagenized with Tn*phoA* (so that background acid phosphatase, encoded by *phoN*, would not interfere with the measurement of fusion protein activity on alteration of the *phoP* locus) and 1800 individual blue colonies with

30   PhoA fusion protein activity were isolated on LB agar plates containing XP. These colonies were the result of 18 separate matings with approximately 20 pools in each. These strains were tested for reduction of fusion protein activity on acquisition of the *pho-24* allele (CS031),

35. which resulted in a PhoP^c phenotype. AP assays were then performed on strains isogenic except for the *phoP* locus.

- 76 -

The PhoP^c phenotype was confirmed in these strains by preparation of whole cell protein extracts and SDS-PAGE analysis. All strains with a PhoP^c phenotype demonstrated the expected distinctive pattern of protein

5   expression in PhoP^c strains, i.e. repressed protein species of specific sizes.

Eight strains were identified with gene fusions to *phoP*-repressed genes. As shown in Table 6, the synthesis of most *prg*::Tn*phoA* fusion proteins was fully repressed

10   by the *pho-24* allele. While two loci had complete repression of fusion protein activity, others demonstrated only partial repression. The expression of *pag* in PhoP^c strains is 5-10 fold less than that observed after bacteria are phagocytosed by macrophages suggesting

15   that the degree of repression of some *prg* loci may be greater when *pag* are maximally activated within acidified macrophage phagosomes.

Lower values for *prgB* -*phoA* fusions in strains with a wildtype *phoP* locus (Table 7B) compared to PhoP$^-$

20   strains (Table 7) may represent some degree of repression in the presence of PhoP.

- 77 -

Table 6

Allele	PhoP⁻	PhoPᶜ	Fold Repression
prgA1::TnphoA	29	7	4
prgB1::TnphoA	137	27	5
prgB2::TnphoA	77	19	4
prgC1::TnphoA	14	1	14
prgE1::TnphoA	21	5	4
prgE2::TnphoA	34	6	6
prgE3::TnphoA	25	6	4
prgH1::TnphoA	92	2	46

- 78 -

In Table 6, a comparison of the effect of *phoP* locus
mutations on Prg-PhoA fusion protein activity is made.
PhoP⁻ indicates that the strain assayed contains the
phoP12 allele (CS030) and PhoP^c indicates the strain
assayed contains the pho-24 allele (CS031). Values were
calculated from stationary phase cultures. The numbers
denote representative values of experiments performed on
three separate occasions and represent activity in units
of AP as defined above.

- 79 -

Table 7A

Strain Allele		Starvation Media	Rich Media
IB010	prgB1::TnphoA	21	26
IB040	prgH1::TnphoA	7	181
CS119	pagC1::TnphoA	1263	102

Table 7B

Strain	Allele	Aerobic	Microaerophilic	Anaerobic
IB010	prgB1::TnphoA	33	777	1521
IB040	prgH1::TnphoA	142	85	41
CS119	pagC1::TnphoA	431	173	81

Table 7C

Strain	Allele	pH 4.5	pH 7.0
IB010	prgB1::TnphoA	332	26
IB040	prgH1::TnphoA	8	18
CS119	pagC1::TnphoA	145	27

- 80 -

        Table 7 demonstrates the effects of environmental
conditions on the *in vitro* regulation of *prg* loci.
Table 7A shows the effect of starvation on *prg* and *pag*
expression.   Starvation medium (N⁻C⁻P⁻) (17) contained 0.04%
5 glucose, 10 mM NH_4Cl, and 1 mM $NaH_2PO_4 \cdot H_2O$. The fusion
protein activity for starvation media was measured after 48
hours of growth (OD_{600} = 0.5) while that in rich media (LB)
was measured in late-logarithmic growth (OD_{600} = 1.0).* All
cultures were grown aerobically.
10   Table 7B shows the effect of oxygen tension on
expression of *phoP*-activated and *phoP*-repressed genes.
Expression in rich medium is compared under aerobic
conditions at stationary phase (OD_{600} > 1.4), microaerophilic
(OD_{600} = 0.8), and strict anaerobic conditions with 80% N_2,
15 10% O_2, and 10% CO_2 (OD_{600} = 0.6) after 24 hours of growth.*
        Table 7C shows the effect of pH on the expression
of fusion protein activity of *prg* and *pag* loci. Expression
was measured from cultures grown to logarithmic growth
(OD_{600} = 0.5) in LB media buffered to various pHs with sodium
20 citrate.   All the numbers represent activity in units of AP
as defined above.

<u>Chromosomal location of *prg*::Tn*phoA* loci</u>

        prg::Tn*phoA* linkage analysis was performed to a
bank of strains with randomly spaced Tn*10*Δ16Δ17 insertions
25 to determine chromosomal locations and whether *prg*::Tn*phoA*
alleles were unlinked loci.   The *prg*::Tn*phoA* insertions were
in five distinct linkage groups.   Three alleles, *prgE1-
3*::Tn*phoA* were identically linked to the Tn*10*Δ16Δ17
insertion of AK3091(26%) and two other alleles, *prgB1-
30 2*::Tn*phoA* were similarly linked to the Tn*10*Δ16Δ17 insertion
of AK3190 (94%), AK3249 (89%), and AK3186 (50%).   Another
allele, *prgH1*::Tn*phoA*, was found to be 37% linked to the
Tn*10*Δ16Δ17 insertion of strain AK3304.   The other two *prg*
alleles did not demonstrate linkage to the bank of strains
35 tested.   The chromosomal DNA of these two strains was

- 81 -

analyzed by Southern hybridization analysis using a portion
of TnphoA as a probe, and a rough physical map of the sites
located adjacent to the TnphoA insertion was determined.
These alleles, prgA and prgC, had different restriction
5   endonuclease sites surrounding the TnphoA insertions. In
addition, the repression of prgA and prgC fusion protein
activity in strains with the pho-24 mutation was different;
prgC was completely repressed, while prgA was only partially
repressed indicating that these loci are different. Thus,
10 five unlinked loci encoding envelope proteins repressed in
the PhoP^c phenotype were identified.

        Though three prg loci were identified that were
linked to transposon insertions, none of the Tn10Δ16Δ17
insertions had a known map location. The physical map
15 location of two of these transposon insertions, AK3249 and
AK3304, was analyzed using XbaI restriction endonuclease
digestion and pulse field gel electrophoresis (PFGE). Since
Tn10Δ16Δ17 contains a single XbaI site, these Tn10Δ16Δ17
insertions can be assigned to a specific XbaI fragment of
20 known map location (Liu et al., 1992, J. Bacteriol.
174:16622). AK3249 was assigned to 28-32 min, while AK3304
was assigned to either end of the 58-70 minute fragment.
Further P22 transduction to known markers in those regions
was performed. The Tn10Δ16Δ17 insertion of strain AK3249
25 and prgB1::TnphoA were found not to be linked to the Tn10
insertion of strain TN3061 (6% linked to dcp), which has a
transposon insertion at 28 min, or to the ompD::Tn5
insertion of strain SH7782 at 32 min. prgH1::TnphoA was
found to be very weakly linked to the srl202::Tn10
30 insertion of strain TT520 (<0.1%) at 59 minutes. These data
indicate that prg are unlinked on the Salmonella chromosome,
consistent with the function of PhoP/PhoQ as global
regulators.

- 82 -

The chromosomal location of Tn*phoA* insertions in
phoP-repressed genes (*prg*::Tn*phoA*) was determined by linkage
analysis to a bank of strains with Tn*10*Δ16Δ17 insertions
(Kukral et al., 1987, J. Bacteriol. <u>169</u>:1787, herein
5 incorporated by reference). Cells with Tn*phoA* insertions
were spread on LB agar plates containing 10 μg/ml
tetracycline and 40 μg/ml XP. Then P22 lysates grown on
strains with Tn*10*Δ16Δ17 insertions were spotted onto plates
with a multiprong inoculator. After overnight inoculation,
10 plates were reviewed for linkage by looking for mixed blue
and white colonies. Linkage was confirmed and quantitated
by carrying out individual transductional crosses between
the Tn*10*Δ16Δ17 containing strains and the strain with the
Tn*phoA* insertion. After selection for the Tn*10*Δ16Δ17
15 encoded tetracycline resistance, strains were scored for
loss of blue color and Tn*phoA* encoded kanamycin resistance.
Some Tn*phoA* strains were found to be linked to Tn*10*Δ16Δ17
strains with no known map location. Two of these Tn*10*Δ16Δ17
insertions were physically mapped using PFGE following *Xba*I
20 restriction endonuclease digestion. Based on physical
mapping, linkage analysis to other transposon insertions by
P22 bacteriophage transduction was determined as necessary.
Chromosomal DNA was prepared as described by
Mekalanos, 1983, Cell <u>35</u>:253, herein incorporated by
25 reference, using Proteinase K instead of Pronase.
Purification of plasmid DNA was performed by standard
methods. Restriction endonuclease digestion was performed
according to the recommendations of the manufacturer (New
England Biolabs). DNA, size fractionated in agarose gels,
30 was transferred to Genescreen Plus membranes (New England
Nuclear/Dupont, Boston, MA) for blot hybridization by the
method of Southern well known in the art. DNA probes were
purified from agarose gels by the freeze-squeeze method
(Tautz et al., 1983, Anal. Biochem. <u>132</u>:14) and
35 radiolabelled with [^{32}P]dCTP by the random primer method

- 83 -

(Feinberg et al., 1983, Anal. Biochem. <u>132</u>:6).

<u>Cloning genes from Tn<i>pho A</i> fusions</u>

        The gene encoding <i>prgH</i> has been cloned using
methods described below. The plasmid, pIB01, containing the
5 <i>prgH</i> gene has been deposited with the American Type Culture
Collection on July 9, 1993 (Rockville, MD) and has received
ATCC designation ATCC 75496. Fig. 5 shows the partial DNA
sequence of <i>prgH</i> (SEQ ID NO: 3). Fig. 9 shows the location
and sequence of the entire <i>prgH</i> gene.

10   The genes described herein which have been
identified by Th<i>phoA</i> insertion can be cloned using methods
known in the art (Beattie et al., 1990, J. Bacteriol.
<u>172</u>:6997). Chromosomal For example, DNA of each strain
containing a <i>prg</i>::Tn<i>phoA</i> gene fusion is digested with a
15 restriction enzyme such as BamH1 which cuts at a single site
in Tn<i>phoA</i> maintaining the fusion junction, <i>phoA</i> sequences
and the <i>neo</i> gene. Similarly, a plasmid such as pUC19 is
digested with the same enzyme. Digested chromosomal and
plasmid DNA are ligated overnight at 15°C and transformed
20 into competent <i>E. coli</i>. Transformations are plated on LB
agar containing ampicillin and kanamycin to select for the
<i>bla</i> gene of pUC19 and the <i>neo</i> gene of TnphoA. The
chromosomal DNA containing the <i>prg</i>::Tn<i>phoA</i> gene fusion can
then be sequenced using standard methodology described
25 above, such as the SEQUENASE® (United States Biochemical)
kit. Universal primer (United States Biochemical)
corresponding to DNA sequences in the plasmid or Tn<i>phoA</i>
primer (5'-AATATCGCCCTGAGCA-3') (<u>SEQ ID NO:4</u>) corresponding
to bases 71 to 86 of Tn<i>phoA</i> can be used as primers.

30   To clone the wild type gene, a fragment of
chromosomal DNA flanking Tn<i>phoA</i> sequences can be used to
screen a cosmid gene bank of wild type <i>Salmonella</i> strain
ATCC 10428 using methods described above for cloning wild
type <i>pagC</i>.

- 84 -

Environmental regulation of _prg_ loci

    Since PhoP/PhoQ are environmentally responsive
regulators, the effects of different growth conditions on
prg::Tn_phoA_ expression were tested. The growth rate of
5 strains with _prg_::Tn_phoA_ insertions was comparable to wild-
type organisms under all conditions. The expression of all
prg loci was maximal in late logarithmic growth phase when
bacteria were grown in rich (LB) media. An example of this
is the comparison of values of _prgH_::Tn_phoA_ expression in
10 Table 7A (rich media and stationary growth) and Table 7C (pH
7.0, log phase). Since the expression of _pag_ loci was
maximal in starvation (which only reaches a maximal OD_{600} =
0.5) and stationary growth phase, this was consistent with a
reciprocal relationship between the expression of _pag_ and
15 _prgs_. Further analysis of _prg_ loci expression under
starvation conditions confirmed this reciprocal relationship
(Table 7A). _prgH_ expression was repressed (Table 7A) and
other _prg_ were minimally affected under starvation
conditions, in contrast to the induction of _pag_ expression
20 when bacteria were starved (Table 7A).

    Because of its role in bacterial-mediated
endocytosis (BME), the effect of oxygen tension in rich
medium on _pag_ and _prg_ expression was also tested (Table 7B).
Different but not reciprocal regulation of _pag_ and _prg_ loci
25 was found on growth at different oxygen tensions. Though
pagA and _pagB_ loci were minimally affected by growth at
different oxygen tensions, the _pagC_ virulence locus was
approximately 5 fold repressed when bacteria were grown
anaerobically as compared to aerobic growth (Table 7B).
30 Variability was also noted in the expression of _prg_ loci in
response to growth conditions in the absence of oxygen. One
loci, _prgH_, was repressed three-fold in anaerobic growth,
while another locus, _prgB,_ was induced almost 50-fold when
grown anaerobically (Table 7B). Other _prg_ loci had minimal
35 change in fusion protein expression as a result of different

- 85 -

oxygen tensions in the growth media.

Low pH conditions also had a variable effect on *prg*
expression (Table 7C). The expression of *pagC* fusion
protein activity was induced under acid conditions as
5 previously known. When bacteria were grown to mid-
logarithmic growth, no significant induction of the relative
repression of *prgH* expression was noted in media of low pH,
while *prgB* expression was induced on exposure of bacteria to
low pH (Table 7C). Hence, loci maximally expressed under
10 diverse environmental conditions can all be repressed by the
PhoP^c phenotype.

Acid sensitivity was tested by the method of Foster
et. al., 1990, J. Bacteriol. 172:771, herein incorporated by
reference. Strains were grown aerobically in E media and
15 0.4% glucose at 37°C to an OD$_{600}$ of 0.5. The pH of the
bacterial culture was decreased to near 3.3 by the addition
of 1 M hydrochloric acid. An aliquot was taken immediately
(t$_o$), the remainder of the culture was incubated further at
37°C with subsequent aliquots removed at 40 min (t$_{40}$) and 80
20 min (t$_{80}$) time points. The pH of the cultures remained near
3.3. The aliquots were diluted 1:10 in cold PBS, washed
and resuspended in normal saline prior to plating serial
dilutions for colony forming units.

prgH is a virulence locus for S. typhimurium

25   Since the PhoP^c phenotype resulted in virulence
attenuation and repressed the synthesis of approximately 20
proteins, the virulence of strains with single mutations in
prg loci was tested (Table 8). Strains with *prg*::Tn*phoA*
insertions were screened for virulence defects by
30 intraperitoneal injection of approximately 150 organisms
into BALB/c mice. Controls were also performed with wild-
type bacteria. A significantly longer time course of
clinical disease progression was observed with a *prg* mutant
strain compared to wild type bacteria. Mice injected
35 intraperitoneally with strains containing the *prgH1*::Tn*phoA*

- 86 -

insertion developed clinical signs of typhoid fever, such as
a "scruffy "phenotype (fever and piloerection) and
hepatosplenomegaly in approximately 10-14 days, compared to
approximately 24 hours for the wild type bacteria. Despite
5 the extended time course of disease development, all the
mice eventually died. Disease progression of mice injected
with other strains containing *prg*::Tn*phoA* insertions showed
a similar pattern of illness to that of wild type bacteria.

- 87 -

Table 8

Intraperitoneal injection			LD_{50}
14028s	Wild type		<10
IB040	*prgH1*		5.6×10^1
CS015	*phoP-102*		6.7×10^5
IB041	*prgH phoP-102*		1.2×10^7

5

Oral inoculation			
14028s	Wild type		6.5×10^4
IB040	*prgH1*		6.5×10^5

- 88 -

Table 8 shows the effect of the *prgH1*::Tn*phoA*
mutation on *Salmonella* mouse virulence. Strains were
isogenic and administered by intraperitoneal injection and
oral inoculation in 35 day old BALB/c mice. The number of
5 animals used at bacterial dilutions near the LD_{50} for each
allele is listed in parentheses. The LD_{50} determinations
were repeated on three separate occasions.

Further testing of the LD_{50} of strains containing
prgH mutations was performed. *prgH* mutants were determined
10 to have an LD_{50} of approximately 60 organisms compared to a
value of <10 for wild type bacteria. Due to the difficulty
in accurately delivering organisms in small doses to mice, a
strain with a mutation in both *prgH* and *phoP* was
constructed. The PrgH⁻PhoP⁻ strain had greater than a 10
15 fold increase in LD_{50} compared to CS015, an isogenic PhoP⁻
strain (Table 8). The combined effect of the two mutations
further documented that the *prgH1*::Tn*phoA* mutation
attenuated *S. typhimurium* virulence and indicated that
mutations which affected two phases of PhoP/PhoQ regulated
20 gene expression were additive in their effect on virulence.
Strains with *prgH1*::Tn*phoA* insertions were also tested for
virulence when administered by the oral route. A 10 fold
decrease in virulence (increase in LD_{50}) was observed (Table
8).

25   Further analysis of the efficiency of strains with
prgH1::Tn*phoA* insertions in crossing the mucosal barrier was
tested by competition experiments with wild-type bacteria.
During the first 72 hours after oral inoculation with mutant
bacteria, no *prgH1*::Tn*phoA* mutants were recovered from the
30 bloodstream of mice compared to control experiments in which
organisms were routinely isolated from the blood of mice
inoculated with wild type bacteria. Other strains with *prg*
mutations were also tested for virulence defects by the oral
route, but no significant change in virulence was observed.

Mouse virulence studies were carried out as
follows. Bacteria were grown aerobically at 37°C to
stationary phase, washed with LB, and diluted in normal
saline. 35 days old (16-18g) female BALB/c mice were
5 purchased from the Charles River Breeding Laboratories, Inc.
(Wilmington, MA). Diluted bacterial samples in saline were
injected intraperitoneally with an inoculum of 0.1-0.15 ml.
Bacteria were administered orally as a 0.5 ml bolus to mice
fasted for 2 hours, via a 2 inch straight, 18 gauge
10 stainless steel animal oral feeding needle (Harvard
Apparatus, Inc., South Natick, MA) under mild 2-bromo-2-
chloro-1,1,1-trifluoroethane (Halothane) anesthesia. The
number of organisms administered was quantitated by plating
for cfu/ml on LB agar. Mouse 50% lethal dose (LD_{50}) values
15 were determined by standard methods (Reed and Muench, 1938,
Amer. J. Hygiene 27:493). The LD_{50} determinations were
repeated on three separate occasions. Competition assays
were performed after bacteria were administered orally to
mice as above. Bacteremia was assessed on days 1-4 from
20 tail bleeds or intracardiac punctures with 50 μl of blood
plated immediately and after growth in LB broth at 37°C
overnight. Spleen and intestinal harvests were performed on
days 1-6 with organs homogenized in 3 mls of 0.9% sodium
chloride. Samples and cultures were plated in serial
25 dilutions. *S. typhimurium* was confirmed by characteristic
growth (black colonies) on Hektoen-enteric agar (Difco
Laboratories) and by the macroscopic slide agglutination
test with *Salmonella* rabbit serum Group B (Antigens 4, 5,
12) (Fisher Scientific).

30 <u>Mutations in oxygen-induced genes do not affect mouse
virulence</u>

Both *prgH* and *pagC* loci were shown to be repressed
by anaerobic growth and required for full virulence, thus
suggesting that a shift from anaerobic to aerobic conditions
35 might serve as a general signal for induction of virulence
genes. Strains with mutations in oxygen-inducible loci

- 90 -

(Aliabadi et al., 1986, J. Bacteriol. <u>165</u>:780) were
constructed. ATCC14028s derivatives with *oxiA*, *oxiC*, and
oxiE mutations were made (termed CS032, CS033, CS034,
respectively). These strains were as virulent as wild type
5 bacteria. Though these gene fusions could still mark
operons containing virulence genes, this data suggests that
these loci are not essential to full virulence and that
oxygen induction is not always correlated with virulence
function.

10 <u>*prgH* mutants have normal survival within macrophages</u>
Since the PhoP^c phenotype resulted in a defect in
bacterial survival within macrophages, the effect of this
mutation on the synthesis of a *prgH*-encoded protein was
tested. A strain with the *prgH1*::Tn*phoA* insertion was
15 tested for intracellular survival within bone marrow-derived
macrophages from BALB/c mice and J774.2 cells, a macrophage
derived cell line. No defect in intracellular survival was
observed. A strain with a *prgB1*::Tn*phoA* insertion was also
tested and found not to have a defect in survival within
20 macrophages.

Assays to determine bacterial survival within
macrophages were performed as described by Buchmeier al.,
1989, Infect. Immun. <u>57</u>:1, herein incorporated by reference.
Bacteria grown to stationary-phase were opsonized for 30
25 minutes in normal mouse serum before exposure to cultured
bone marrow-derived macrophages harvested from BALB/c mice.
One hour after infection, gentamicin 10 μg/ml was added to
kill extracellular bacteria. All time points (1, 4, and 24
hr) were done in triplicate and repeated on three separate
30 occasions.

Cultured bone marrow macrophages were harvested
from BALB/c mice purchased from the Charles River Breeding
Laboratories. J774.2 macrophages were cultured in
Dulbecco's minimal essential medium with 10% fetal bovine
35 serum (DMEM/10%FBS).

- 91 -

prg::Tn*phoA* insertions do not suppress the phenotypes of
PhoP mutants

Several phenotypes of *phoP* mutants, including
defensin and acid sensitivity as well as mouse virulence
5 attenuation, were tested for suppression on addition of a
prg::Tn*phoA* mutation. To test the ability of a *phoP*
mutation to suppress the synthesis of *prg* products, PhoP
mutant strains isogenic except for *prg*::Tn*phoA* mutations
were constructed and tested for mouse virulence, where
10 suppression would involve an increase in virulence, or
decreased acid and defensin sensitivity. *prg*::Tn*phoA*
insertions had no effect on the virulence phenotypes of
PhoP⁻ bacteria. These results indicate that the *prg*::Tn*phoA*
mutations tested did not suppress the PhoP null phenotype as
15 single mutations.

PrgH and PhoPᶜ mutants are defective in bacterial-mediated
endocytosis by cultured epithelial cells

The BME of *prg*::Tn*phoA* and PhoPᶜ *S. typhimurium*
strains was tested. The following observations (described
20 herein) suggested that *prg* genes may be involved in
bacterial-mediated uptake by eucaryotic cells:
prgH1::Tn*phoA* was shown to be located at 59' on the
bacterial chromosome, a location where other genes essential
to invasion are clustered; *prgH* mutants were shown to be
25 defective in competition with wild type organisms on
reaching the bloodstream of mice in the first 72 hours after
oral ingestion; and the expression of one *prg* locus, *prgB*,
was dramatically induced under anaerobic growth conditions.
Strains with *prgH* and *pho-24* mutations had a significant
30 reduction (p-value < 0.01) in their ability to induce uptake
by Madin-Darby canine kidney (MDCK) polarized epithelial
cells compared to wild-type bacteria. Other *prg* strains
with Tn*phoA* insertions did not demonstrate a statistically
significant defect in BME by epithelial cells (Table 9).
35 The adherence of strains defective in BME was unaffected by
the *prgH*::Tn*phoA* insertion when determined by cell-

- 92 -

associated cfu/ml before the administration of gentamicin
(Table 9) and by microscopy.

         . To assay bacterial adherence and uptake of bacteria
by epithelial cells, bacterial strains were grown at 37°C
5 without shaking (microaerophilic) to a final density of
approximately $2x10^8$ colony forming units (cfu)/ml.   Assays
were performed by seeding 10^5 MDCK cells/well in 24-multiwell
tissue culture plates.   Cells were incubated overnight at
37°C in 5% CO_2/95% air atmosphere in DMEM/10%FBS without
10 antibiotics until >80% confluent.   The adherence and
invasion assays were carried out according to the protocol
of Lee and Falkow, 1990, Proc. Natl. Acad. Sci. USA 87:4304,
herein incorporated by reference.

- 93 -

Table 9

Strain	Genotype	Adherence	Invasion
14028s	Wild type	4.2%	3.8%
SM7	Str^r smb	---	0.6%*
CS119	pagC1::TnphoA	---	1.9%
IB005	prgA1::TnphoA	---	7.6%
IB010	prgB1::TnphoA	---	2.9%
IB020	prgC1::TnphoA	---	1.5%
IB025	prgE1::TnphoA	---	1.9%
IB040	prgH1::TnphoA	5.7%	0.1%*
CS022	pho-24	1.9%	0.06%*
IB043	pWKSH5 in IB040	---	17.5%*
IB044	pWKSH5 in CS022	---	0.09%*

- 94 -

In Table 9, the effect of *prg*::Tn*phoA* insertions
on *Salmonella*-mediated endocytosis by MDCK epithelial
cells is shown. Microaerophilically grown bacterial
strains were assessed for changes in adherence and
5   invasion. Adherence was determined as the percentage of
bacteria adhered to the cells after centrifugation and 30
minute 4°C incubation/ total number of bacteria added to
each well. Invasion was determined as the percentage of
bacteria that had invaded after a two hour incubation
10   with gentamicin/ total number of bacteria added to each
well. There was no difference between *S. typhimurium*
wildtype and *S. enteritidis* CDC5 wildtype strains with
respect to adherence and invasion frequency. The
asterisk (*) represents statistical significance by
15   variance analysis of the invasion data done in triplicate
compared to wild-type (p-value < 0.01).

The confluent MDCK monolayers were washed three
times with PBS, then 0.9 ml of cold DMEM/10%FBS was added
to each well. Bacteria were washed in LB and resuspended
20   in an equivalent volume of DMEM/10%FBS. Approximately
5×10^7 bacteria were added/well. The plates were spun at
500 rpm at 4°C for 10 minutes, then incubated at 4°C for
30 minutes. Adherent bacteria were recovered by washing
the plates three times with phosphate-buffered saline
25   (PBS), lysing the epithelial cells in 0.5 ml of 1%
Triton-X-100/PBS, and plating for cfu/ml on LB agar. A
morphologic assessment of adherence was also performed by
staining bacterially infected epithelial cell monolayers
grown overnight on coverslips for 7 minutes in 1 μg/ml 4'
30   6-diamidino-2-phenylindole (DAPI). These DAPI stained
coverslips were examined by both fluorescent and phase
contrast microscopy using a Leitz Laborlux 12 microscope.

Invasion or bacterial-mediated endocytosis (BME)
was assessed by allowing bacteria to adhere as described
35   above. Plates containing bacteria and epithelial cells

- 95 -

were incubated for two hours at 37°C in a 5% CO_2/95% air
atmosphere. Each well was washed three times with PBS to
remove bacteria not associated with cells. DMEM/10%FBS
supplemented with 10 μg/ml gentamicin was then added to
5 kill extracellular bacteria. After 90 minutes of
incubation, the cell monolayers were washed three times
with PBS and the viable intracellular bacteria were
released by vigorously pipetting with 0.5 ml of 1% Triton
X-100/PBS. An invasion deficient *Salmonella enteritidis*
10 mutant and an invasive clinical wild-type isolate of *S.
enteritidis* were used as controls for BME. Viable
bacteria were quantitated by plating for cfu/ml on LB
agar medium. All assays were done in triplicate and
repeated at least three times.

15 MDCK epithelial cells were used between passage
40-58 to maximize bacterial adherence and invasion.
Epithelial cell lines were cultured in DMEM/10% FBS and
1% penicillin/streptomycin solution at 37°C in a 5% CO_2
atmosphere.

20 To assay bacterial defensin sensitivity, NP-1
defensin was purified from rabbit peritoneal neutrophils
according to methods known in the art (Selsted et al.,
1985, J. Biol. Chem. 260:4579; Selsted et al., 1984,
Infect. Immun. 45:655). Typically, 10^5 bacteria in 0.5%
25 tryptone in 100 μl volume were exposed to 50-100 μg of
defensin/ml at 37°C for 2 hours. The reactions were
stopped by diluting the reaction in 0.9% NaCl.
Appropriate dilutions were plated to determine the cfu/ml
of surviving bacteria. Assays were performed in
30 duplicate at least twice for each strain. Appropriate
assays with sensitive (PhoP⁻) and resistant (wild-type)
strains were performed as controls.

Mapping of *prgH*

The location of *prgH* relative to other invasion
loci at 59 minutes was determined using linkage analysis.
P22 transduction linkage analysis indicated that the
Tn10Δ16Δ17 of strain AK3304 had similar linkage to *invA*
5   (40%) and *prgH* (37%); however, *invA* was not linked to
sorbital. The *prgH1*::Tn*phoA* insertion was found to be
linked (99.6%) to the transposon insertion of EE517, a
strain with a 8.5 kilobase deletion adjacent to the
Tn5B50- 378 insertion of *hil*.

10   A physical map of the restriction endonuclease
sites surrounding the Tn*phoA* insertion of strain IB037
was made (Fig. 4) revealing no similarities to the known
restriction endonuclease map of the *invA-E* region.
Plasmids containing the cloned *inv* and *hil* DNA were then
15   used as probes in Southern hybridization analysis of
chromosomal DNA from wild type ATCC10428s and IB040
bacteria containing the *prgH1*::Tn*phoA* insertion. When a
plasmid which contains other invasion loci highly linked
to *invA-E* (*invH*, *invF*, and part of *invG*) was used as a
20   probe, no differences in hybridization pattern was found
between wild type bacteria and strain IB040 indicating
that *prgH* was not located within the *inv* region.
However, when a plasmid containing a 5 kb region
immediately downstream of the Tn5B50-380 insertion of
25   *hil* was used as a probe, the *prgH1*::Tn*phoA* insertion was
demonstrated to be located within this region. By use of
the known restriction map of the *hil* locus (Lee et al.,
1992, Proc. Natl. Acad. Sci. USA <u>89</u>:1847) and the known
restriction endonuclease sites of Tn*phoA*, the physical
30   map of this area and the relationship of *prgH1*::Tn*phoA*
within it were further defined (Fig. 4). The
prgH1::Tn*phoA* insertion was oriented so that the
direction of transcription of the *phoA* fusion protein was
opposite to that of the Tn5B50 insertions that confer the
35   *hil* phenotype and contain a constitutive neomycin

promoter that is transcribed out of the transposon (Fig.
4). Although *prgH* was found to be located within the *hil*
locus, this gene is unique in that it is oppositely
transcribed and unlike any other genes identified within
5   the *hil* locus, *prgH* is regulated by the *phoP* regulon.

Since it was possible that a protein whose
expression was altered by the Tn5B50-380 insertion might
alter the expression of *prgH*, strains containing both
insertions were constructed and the *prgH-phoA* fusion
10   protein activity compared under different environmental
conditions. When bacteria were starved or grown
anaerobically, derepression of fusion protein activity
was observed. Table 11 shows the effect of the Tn5B50-
380 insertion on expression of *prgH* fusion protein
15   activity.

Table 11

Strain	Allele	Starvation	LB (aerobic)	LB(anaerobic)
IB040	*prgH1::TnphoA*	5	142	41
IB042	Tn5B50-380 *prgH1::TnphoA*	46	248	227

This data demonstrates that the Tn5B50-380
insertion increased *prgH* expression, even though *prgH*
transcription was opposite to that of the Tn5B50-380
encoded neomycin promoter. Starvation (repressing
25   conditions for *prg*) indicates that bacteria were grown
aerobically for 48 hours in starvation medium (N⁻C⁻P⁻)
containing 0.04% glucose, 10 mM NH_4Cl, and 1 mM
$NaH_2PO_4 \cdot H_2O$. LB (aerobic) indicates that bacteria were
grown in Luria-Bertani broth (rich media) to late
30   logarithmic growth (nonrepressing conditions) (OD_{600}
>1.0). LB (anaerobic) indicates that bacteria were grown
under strict anaerobic conditions for 24 hours (OD_{600} =
0.6). All the numbers represent activity in units of AP
as described above.

- 98 -

    To rule out the possibility that the BME defect of
the *prgH* mutant was an artifact of the PhoA fusion
protein produced, complementation analysis was performed
with a plasmid (pWKSH5) containing a 5.1 kb *HindIII*
5   fragment which included the *hil* and *prgH* loci. The
plasmid was crossed into PrgH (IB040) and PhoP^c (CS022)
mutant bacteria to create strains IB043 and IB044,
respectively. The BME phenotype of the PrgH mutant was
similar to wild-type with the same plasmid insertion.
10   The BME phenotype of the PhoP^c mutant was not complemented
by this plasmid. These results indicate that a gene
product altered in synthesis as a result of the
prgH::Tn*phoA* insertion was necessary for BME.
    Using a strain with a *phoP/phoQ* locus mutation
15   that constitutively simulates the environmental
activation of *pag* (phenotype PhoP^c), five unique *phoP-*
repressed loci encoding envelope proteins were defined.
phoP-repressed genes (*prg*) were found to be widely spaced
on the chromosome and the expression of *prg* loci was
20   repressed under starvation conditions, when *pag* loci
were induced (Table 10).

<u>Table 10</u>

<u>Environment</u>	<u>*pag*</u>	<u>*prg*</u>
media	starvation	rich
O$_2$	aerobic - *pagC*	aerobic -*prgH*
	anaerobic - *prgB*	
pH	3.3-5.5	3.3-5.5 -*prgB*
		>6.0 - *prgH*
mammalian cell	macrophage	epithelial

    PrgH was shown to lie between two Tn5B50 insertions
that confer the *hil* phenotype. Since deletion mutants in
this region have been demonstrated to also have defects
of BME, and the BME defect of *prgH* mutants can be
35   complemented with a plasmid containing this locus, it is

possible that a protein not synthesized as a result of
the *prgH1::TnphoA* insertion promotes BME (Fig. 4).

Contrary to the expectation that genes essential
to the *hil* phenotype would be induced under
5 microaerophilic conditions similar to what was found for
prgB, *prgH* expression was maximal during aerobic growth
and the Tn5B50-380 insertion, which results in a *hil*
phenotype, derepressed expression of *prgH*. In addition,
the direction of transcription predicted by the
10 *prgH1::TnphoA* insertion is opposite to that of the
Tn5B50-380 encoded neomycin promoter associated with the
hil phenotype suggesting that a regulatory protein
interrupted by or transcribed from the Tn5B50-380
insertion affects the expression of *prgH*.

15   In view of the observation that pWKSH5, a plasmid
containing *prgH* (*hil*), did not complement PhoP^c bacteria
for BME, it is possible that other invasion genes may
also be regulated by PhoP/PhoQ. If *prgH* was expressed
from pWKSH5, despite the presence of the *pho-24* mutation,
20 this suggest that other genes repressed as part of the
PhoP^c phenotype are necessary for BME.

The identification and characterization of *prgH*
has shown that PhoP/PhoQ oppositely regulate factors
necessary for bacteria to enter or to survive within
25 mammalian cells, further documenting the importance of
gene regulation to bacterial virulence. The
identification of *prg* loci can be used to study the
regulation of bacterial genes after infection of
mammalian cells. Understanding the regulation of
30 virulence genes, such as *prgH* can also be used to
attenuated pathogenic bacteria for the development of new
live vaccines for typhoid fever.

Role of *prg* genes in virulence

The *prg* locus, *prgH*, was found to contribute to
35 mouse virulence when *S. typhimurium* was administered by

- 100 -

both the oral and intraperitoneal routes. PrgH as well
as PhoP^c mutants were further found to be defective in
bacterial-mediated uptake by epithelial cells suggesting
that an inability to cross epithelial barriers might

5   contribute to the attenuation of virulence observed.
Competition studies in mice after oral ingestion of
bacteria further supported that *prgH* mutants were
defective in transcytosis across the intestinal
epithelial barrier. Therefore, at least two phases of

10   PhoP/PhoQ regulated protein expression essential to
bacterial virulence have been defined. In one phase, *prg*
expression promotes bacterial mediated endocytosis by
epithelial cells (Table 10), while in another phase, *pag*
expression promotes survival within macrophages.

15   Systemic pathogens, such as *Salmonella*, may
encounter more complex and varied environments than may
be encountered by mucosal pathogens. The achievement of
intermediate states of *pag* and *prg* expression could be
essential to virulence at some stage of the infectious

20   cycle. Consistent with this concept was the lack of
uniformity observed in the expression of *pag* and *prg* on
growth at different oxygen tensions and pH conditions.
These data may also indicate that not all regulation of
pag and *prg* is mediated directly through PhoP and PhoQ.

25   Given the function of PhoP as a transcriptional
regulator, it is likely that *prg* loci repression occurs
at the level of transcription.

          The approach of defining genes repressed by the
pho-24 mutation has led to the discovery of at least one

30   virulence locus, *prgH*, which can be mutated to attentuate
the bacteria for vaccine purposes.

Attenuation of Bacterial Virulence by Constitutive
Expression of Two-component Regulatory Systems

          The virulence of a bacterium can be attenuated by

35   inducing a mutation which results in the constitutive
expression of genes under the control of a two-component

regulatory system or by inducing a mutation that
inactivates a gene under the control of the two-component
systems. A balance between the expression of the genes
under the control of the two-component system, e.g.,
5 between *pag* and *prg* gene expression, and possibly between
two-component system regulated genes and other genes, is
necessary for full virulence. Mutations that disrupt
this balance, e.g., mutations that cause the constitutive
expression of a gene under the control of the two-
10 component system, or a mutation that inactivates a gene
under the control of the two-component system, e.g., the
pag gene, reduce virulence.

Constitutive mutations in two-component regulators
can be identified by the use of a strain containing a
15 recorder gene fusion to a gene regulated by the two-
component system. Such gene fusions would most typically
include DNA encoding the *lacZ* gene or AP fused to a gene
under the control of the two-component system. Strains
containing fusions that are (as compared to wild type or
20 parental strains) highly expressed in an unregulated
fashion, i.e., constitutive, can be detected by increased
color on chromogenic substrates for the enzymes. To
detect constitutive mutations a cloned virulence
regulator could be mutagenized e.g., by passage through
25 an *E. coli* strain defective in DNA repair or by chemical
mutagenesis. The mutated DNA for the regulator would
then be transferred to the strain containing the gene
fusion and constitutive mutations identified by the high
gene fusion expression (blue color in the case of a *lacZ*
30 fusion grown on media containing X-gal). Constitutive
mutations in a component of a two-component regulatory
system could also be made by *in vitro* mutagenesis after
other constitutive mutations have been sequenced and a
specific amino acid change responsible for the
35 constitutive phenotype identified. Putting several amino

- 102 -

acid changes that all result in a PhoP constitutive
phenotype would result in a decreased frequency of
reversion by spontaneous base changes. A constitutive
mutation could also be constructed by deletion of the
5   portion of the amino terminus of the phospho-accepting
regulator which contains the phosphoacceptor domain e.g.,
deletion of sequences encoding amino acids amino terminal
to amino acid 119 in the *phoP* gene or deletion of
analogous phospho accepting sequences in genes of other
10   two-component regulatory systems. This could result in a
conformational change similar to that induced by
phosphorylation and result in increased DNA binding and
transcriptional activation.

Attenuation of virulence: deletion in the phoP/phoQ
15   regulon

As discussed above, the PhoP regulon is essential
to full virulence of *Salmonella*. This regulon is
composed of two genes, PhoP and PhoQ located in an
operon, and over 40 genes they positively and negatively
20   regulate (pag and prg, respectively).

PhoP null *S. typhimurium* mutants have been
demonstrated to be markedly attenuated and also effective
vaccine strains when studied in the BALB/c mouse model of
typhoid fever. This phenotype is likely the result of
25   multiple, phoP-activated virulence genes, as transposon
insertions in multiple different phoP-activated genes
have been independently demonstrated to decrease *S.
typhimurium* virulence. *S. typhimurium* mutants deleted
for genes essential to aromatic amino acids (aroA null or
30   aroC/aroD null mutants) are also markedly attenuated in
the mouse model. However, testing of *aroC/aroD* mutants
in humans has shown that although these strains are
immunogenic, bacteremias and side effects such as fever
have been noted at doses as low as 10^5 to 10^7 organisms
35   administered as a single oral dose (J. Clin. Invest.
90:412-420).

- 103 -

It has now been found that a large deletion in a
global regulator of *Salmonella* virulence, i.e., the
PhoP/PhoQ operon, significantly decreases the virulence
of the bacteria. This mutation, the result of a 1 kB
5   deletion of DNA within the *phoP/phoQ* locus, was initially
made in *S. typhimurium* and subsequently transferred via
homologous recombination to *S. typhi*. In order to confer
an even greater margin of safety in construction of this
vaccine, it was created in a strain background deleted
10   for genes essential to aromatic amino acids and carrying
the histidine G46 mutation, a mutation rendering the
organism auxotrophic for histidine. The resulting
strain, *S. typhi* TyLH445, offers several advantages over
existing vaccine candidates, most notably, immunogenicity
15   without transient bactermia.
Use
        The *Salmonella* cells of the invention are useful
as sources of immunological protection against diseases,
e.g., typhoid fever and related diseases, in an animal,
20   e.g., a mammal, e.g., a human, in particular as the basis
of a live-cell vaccine capable of colonizing the
inoculated animal's intestine and provoking a strong
immune reaction. Appropriate dosages and conditions of
administration of such a live, attenuated vaccine are
25   known in the art, e.g., as described in Holem et al.,
Acute Enteric Infections in Children, New Prospects for
Treatment and Prevention (1981) Elsevier/North-Holland
biomedical Press, Ch. 26, pp. 443 et seq. (Levine et
al.), hereby incorporated by reference, and are described
30   in the examples below.
Advantages
        One advantage of the invention is that the
bacterial cells are attenuated as a result of a
mutation(s), i.e., the *phoP/phoQ* operon, that directly
35   affect a virulence pathway. Another advantage is that

- 104 -

the bacterial cells have mutations in two completely different attenuating genes, i.e., the aromatic amino acid synthesis pathway (*Aro*), and in an operon important to *Salmonella* virulence (*PhoP/Q*). As a result, the

5   bacteria appear to be extremely attenuated; doses as high as 1×10^9 cfu appear to be very safe. Other vaccines under development, such as CVD 908, have caused some systemic symptoms, e.g., fever or bacteremia, at doses as low as 1×10^7 cfu.

10   In addition to the *phoP/phoQ* deletion and the AroA-mutation, the bacterial cells of the invention may also contain a histidine mutation to further alternate virulence, although absence of the histidine mutation may improve immunogenicity. The bacterial cells of the

15   invention are the most promising vaccine candidates to date because they are strongly immunogenic and safe, i.e., extremely attenuated.

EXAMPLE 1: Construction of vaccine strain

        The bacterial cells of the invention were made by

20   deleting approximately 1 kb of DNA in the *phoP/phoQ* regulon.

        PhoP/phoQ deleted suicide vectors were constructed using methods known in the art. A DNA fragment containing the *phoP/phoQ* locus was obtained by PCR using

25   wild type *S. typhimurium* chromosomal DNA as a template. PCR primers flanking the *phoP/phoQ* locus were engineered to contain terminal restriction enzyme recognition sites, e.g., recognition site for EcoRI, to facilitate subsequent cloning. Following amplification, the PCR

30   product was digested with EcoRI and cloned into the EcoRI site in the polylinker of a high copy vector. The plasmid containing the *phoP/phoQ* DNA fragment was named pLH356.

        Sequence analysis and restriction mapping of the

- 105 -

phoP/phoQ locus revealed four HpaI sites within the
locus; no HpaI sites were found in the vector. To create
an internal deletion within the *phoP/phoQ* locus, pLH356
DNA was cut to completion with HpaI, and religated, to
5   yield with an internal deletion from nucleotides 376-1322
(pLH418). This deletion was confirmed by restriction
digestion of the plasmid.

A DNA fragment containing the internally deleted
phoP/phoQ locus was excised from pLH418 using the
10 SacI/SphI restriction sites within the polylinker region
of the vector. This fragment was cloned into compatible
sites in the plasmid CVD442, which carries the *sacB* gene
to allow positive selection for allelic exchange. The
resulting suicide vector was called pLH423.

15 pLH423 was transformed into *E. coli lambda pir*
SY327, and subsequently into *E. coli lambda pir* SM10
(strain LH425). *E. coli* strain LH425 was mated with
S. typhimurium strain CS019. Single recombinants
carrying plasmid sequences integrated onto the
20 *S. typhimurium* chromosome were selected by plating on
agar containing ampicillin and chloramphenicol (Strain
LH428). These strains were confirmed to be ampicillin
resistant and sucrose sensitive, i.e., death on 20%
sucrose plates containing no NaCl when incubated at 30°C.
25 These data confirm the integration of plasmid sequences
into the *Salmonella* chromosome.

A P22 bacteriophage lysate was made from strain
LH428; phage particles were concentrated 20x by high
speed centrifugation and transduced into *S. typhi* strain
30 522Ty2 (a strain with a deletion in the *aroA* gene, and
the G646 mutation which renders the organism auxotrophic
for histidine). Single recombinant *S. typhi* organisms
were selected by plating on LB plates supplemented with
aromatic amino acids, cystine, histidine, and ampicillin
35 (strain LH435).

- 106 -

Strain LH453 was grown with aromatic amino acids,
cystine, and histidine (but without ampicillin) to mid
logarithmic growth phase. Serial dilutions were plated
on LB 20% sucrose plates lacking NaCl, and on LB plates

5   lacking NaCl. The number of bacteria that grew on plates
without sucrose was greater than the number that grew on
sucrose-supplemented plates by a factor of three logs.
These data suggest that many colonies lost plasmid
sequences containing the sacB gene.

10   Multiple colonies from the sucrose selection were
picked and confirmed to be ampicillin sensitive and
sucrose resistant. Chromosomal DNA from approximately 10
colonies was purified and subjected to Southern blot
analysis, utilizing the 2.3 kb fragment of wild type

15   phoP/phoQ as a probe.
         Southern blotting revealed the loss of two HpaI
sites and an XmnI site known to be within the 1 kb
deleted fragment of phoP/phoQ in several strains. One of
these strains was designated TyLH445.

20   EXAMPLE 2:   in vitro evaluation of TyLH445
         TyLH445 was extensively characterized in vitro
using standard clinical microbiological tests. The
nutritional requirements of TyLH445 were evaluated.
TyLH445 did not grow on M-9 plates unless supplemented

25   with aromatic amino acid mix, cystine (S. typhi grows
better with cystine), and histidine. These data
confirmed that TyLH445 was AroA-, His-.
         TyLH445 was found to agglutinate with polyclonal
serum against Salmonella and polyclonal serum against S.

30   typhi Vi antigen. Group D agglutination was found to be
variable, perhaps due to excess Vi antigen. TyLH445 was
also found to be indole negative (as are all
Salmonellae), and to produce very little hydrogen sulfide
(as do many S. typhi). Biochemical testing utilizing

- 107 -

both the VITEK system as well as the BBL Crystal Enteric
organism identification system was also carried out.
These data indicated that the TyLH445 strain was *S.
typhi.*

5   Growth characteristics of TyLH445 were also
evaluated. TyLH445 was found to grow just as quickly as
its parent, 522Ty2, (*phoP/phoQ* locus intact). Growth *in
vitro* was measured in aromatic amino
acid/histidine/cystine-supplemented Luria broth at 37°C.
10   Growth curves of the parent and vaccine strain were found
to be essentially identical (see Fig. 10).

        Standardized clinical testing methods were use to
determine antibiotic sensitivity. TyLH445 and the parent
strain, 522Ty2, were found to be sensitive to ampicillin,
15   trimethoprim-sulfamethoxazole, ciprofloxacin,
aminoglycosides, and third generation cephalosporins. No
difference in zone sizes was detected between the parent
and vaccine strains, suggesting that no other antibiotic
resistance mechanisms, e.g., modification of antibiotic
20   transport systems, or modification of the cell wall of
the bacterium, were affected by introduction of the
mutated *phoP/phoQ* locus into *S. typhi.*

        The *phoP/phoQ* HpaI deletion mutants were tested
for defensin sensitivity, a phenotype of *PhoP* null
25   mutants. Defensin sensitivity assays were performed as
follows.

        Liquid cultures of strains to be tested were grown
overnight. Cultures were then diluted 1:200, and grown
to an optical density (OD_{600}) of approximately 0.2, after
30   which the cells were diluted to concentration of
approximately 1×10^5 organisms per 0.05 ml.

        Two reactions were carried out for each strain:
(1) vehicle alone (0.01% acetic acid in sterile water)
and (2) defensin NP-1 solution (70 ug/ml in 0.01% acetic
35   acid). An equal volume of bacterial suspension in

- 108 -

tryptone was added and the test tubes were incubated on a
roller at 37°C for 2 hours. The final volume in each
reaction tube was 0.1 ml, making the final concentration
of defensin 35 ug/ml.

5   Defensin is inactivated by the high salt and high
protein concentration present in bacterial growth media,
e.g. LB broth. Thus, defensin activity was stopped by
adding 900 ul of Luria broth to each tube. Serial
dilutions of each tube were plated and cfu/ml was
10   determined for both the control tube and treatment tube
for each strain. Results were expressed as log of
bacteria killed for each strain. Typically, 1.0-1.5 log
of wild type bacteria were killed. *PhoP* null mutants
generally exhibit 2-4 logs of killing. Since strains
15   with slower growth rates appear less susceptible to
defensin killing, the growth rate of each strain tested
in the defensin sensitivity assay was measured. Strains
with similar growth rates were compared in the defensin
sensitivity assay.

20   The HpaI deletion was evaluated both in an *S.
typhimurium* background and in the *S. typhi* background.
In both backgrounds, the deletion mutation conferred
sensitivity to rabbit defensin NP-1 at a concentration of
35 ug/ml. See Fig. 11 and Fig. 13. The difference
25   between *PhoP+* and HpaI deleted *PhoP* null mutants was less
pronounced in the *S. typhi* strain, an effect that may
reflect the slower growth rate of the less hardy *S. typhi*
strain compared to the *S. typhimurium* strain which lacks
the additional auxotrophies.

30   The state of *phoP* activation in bacteria with the
HpaI *phoP/phoQ* deletion was tested utilizing a *LacZ*
recorder gene fused to *phoP*-activated gene B (*pagB*).
Since the efficiency of transduction utilizing P22 in
S. typhi is low, these studies were performed in *S.
35   typhimurium* rather than *S. typhi*. *PhoP* activation was

- 109 -

found to be 40-60 Miller units (Miller et al., 1972,
Experiments in Molecular Genetics, Cold Spring Harbor
Laboratory, Cold Spring Harbor, N.Y., pp. 352-355) in the
presence of an intact _phoP/phoQ_ locus, and just barely

5   detectable in strains with the HpaI deletion (3cfusee
Fig. 12).

EXAMPLE 3:   _In vivo_ evaluation of _S. typhimurium_ HpaI
deleted strain

    As _S. typhi_ strains are not pathogenic for mice,

10   the HpaI _phoP/phoQ_ deletion mutation was evaluated in
both wild type and _aroA- S. typhimurium_. Female BALB/c
mice were injected intraperitoneally with various
dilutions of _S. typhimurium_ LH430, a wild type _S._
typhimurium carrying the HpaI deletion. The LD_{50} of this

15   strain was determined to be between 8.2×10^5 and $8.2 \times$
10^6. (All mice receiving 8.2×10^5 cfu survived, and all
receiving 8.2×10^6 died.) These data are consistent
with the LD_{50} data obtained with strains harboring
transposon insertions at the _phoP/phoQ_ locus.

20   Immunogenicity of the HpaI _phoP/phoQ_ deletion was
evaluated in _S. typhimurium aroA::tet_ (LH481), a strain
comparable to the human vaccine strain. Mice were
inoculated intraperitoneally with 2.3×10^5 and 2.3×10^6
cfu of LH481 (4 mice per vaccine dose), and challenged 30

25   days later with 30 × the LD_{50} of wild type organisms. All
mice but one mouse survived. The mouse that died was in
the group that received the lower vaccine dose. No
animal receiving the higher vaccine dose became ill.

EXAMPLE 4:   Phase I study human studies

30   The vaccine strain was administered to human
volunteers at doses of 1×10^5 to 1×10^{10} cfu/single oral
dose. Two volunteers received each dose; 3 volunteers
were given a dose of 1×10^8 cfu/ml. Volunteers were
evaluated at various time points following administration

35   of the vaccine.

- 110 -

Safety

       To detect the presence of the vaccine strain in
patient blood, Bactec blood cultures were performed in
duplicate on days 4, 6, 8, 10, 12 after taking vaccine.
5   Bacteremia was not detected in any of the volunteers.
       Thirteen adult human volunteers have received
escalating single oral doses of this new attenuated
typhoid fever vaccine. No individuals have had side
effects of any sort. Specifically, there have been no
10   fevers, no gastrointestinal symptoms, and no
constitutional symptoms. Volunteers have been subjected
to serial blood cultures on a preset schedule after
receiving the oral vaccine 2 sets of BACTEC blood
cultures performed on each of days 4, 6, 8, 10 and 12
15   after receiving the vaccine, and no positive blood
cultures have been noted. Volunteers have been followed
up at 2 months after receiving the vaccine, and no late
symptoms have been reported.

Colonization

20   Stool samples were tested for the presence of the
vaccine strain TyLH455 using methods known in the art.
Primary stool was evaluated for the presence of the
vaccine strain on culture plates. In some cases, it was
necessary to enrich stool samples for the vaccine strain
25   by incubating the stool overnight in BBL Selenite F broth
supplemented with Aro/His/Cystine in order to detect the
bacteria. This medium is somewhat inhibitory for *E. coli*
and but promotes *Salmonella* growth.
       Volunteers have been colonized for various time
30   periods from 1-6 days after receiving the vaccine. With
the highest doses (10^9 or 10^{10}) volunteers have had
positive primary culture plates in the initial 1-3 days
post vaccination, whereas at lower doses, only selenite
enrichment broth cultures (selective medium for
35   *Salmonella* which inhibits other enterics) have been

- 111 -

positive for the vaccine organism. No volunteer studied
thus far has had prolonged carriage of the vaccine
organism at 2 months of followup.

<u>Table 17</u>

Dose	Number	Colonization
10^5	2	NO
10^6	2	2/2 for 1-2 days
10^7	2	1/2 for 3 days
10^8	3	1/3 for 6 days
10^9	2	2/2 for 4-6 days both had positive primary plates day 1
10^{10}**	2	2/2 for 3-6 days both had positive primary plates on days 1 and 2

*   Measured by whole cell and LPS ELISAs and Widal
test vs. H flagellar antigen. Sera analyzed at 1:40 and
higher dilutions in all tests.

**   One of these volunteers has received a booster
dose of 10^{10} organisms, given one month after the primary
inoculation (serologies pending).

<u>Immunogenicity</u>

        Induction of an immune response to the vaccine
strain was measured by standard ELISA assays. Sera was
collected from volunteers 0, 7, 14, 21, and 28 days after
receiving a single oral dose of the vaccine. ELISA
assays were carried out using whole bacteria TyLH445 and
S. typhi LPS (SIGMA, St. Louis, MO) as antigens. Day 0
serum from each volunteer was used as an internal
negative control. Convalescent sera from patients
previously infected with wild type S. typhi (most from

- 112 -

Mexico) were used as positive controls.

Several volunteers had documented seroconversion at 21 days after receiving the vaccine, as measured by ELISA in which IgG antibodies directed against whole
5   vaccine organisms or against *S. typhi* LPS were detected. Sera taken from patients prior to administration of the vaccine (pre-immune sera) were tested, and the data used to establish a baseline. Patient sera taken at various time points after vaccination were considered positive if
10   the test results were 0.2 ELISA OD units greater than that of the preimmune serum.

Other Embodiments

Other embodiments, e.g., strains of *Salmonella* which contain only a deletion in the phoP/phoQ regulatory
15   locus to attenuate virulence, and strains which, in addition to a phoP related mutation or genetic alteration, also contain an attenuating mutation in another gene, e.g., cya gene (adenylate cyclase) or crp gene (adenylate cyclase receptor), are also within the
20   claims.

- 113 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Miller, Samuel I.
                     Mekalanos, John J.

    (ii) TITLE OF INVENTION: SALMONELLA VACCINES

    (iii) NUMBER OF SEQUENCES: 15

    (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: Fish & Richardson
        (B) STREET: 225 Franklin Street
        (C) CITY: Boston
        (D) STATE: Massachusetts
        (E) COUNTRY: U.S.A.
        (F) ZIP: 02110-2804

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version
                          #1.25 and WordPerfect (Version
                          5.1)

    (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER:
        (B) FILING DATE:
        (C) CLASSIFICATION:

    (vii) PRIOR APPLICATION DATA:

        (A) APPLICATION NUMBER: 08/090,526
        (B) FILING DATE: July 9, 1993

    (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME: Clark, Paul T.
        (B) REGISTRATION NUMBER: 30,162
        (C) REFERENCE/DOCKET NUMBER: 00786/220001

    (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: (617) 542-5070
        (B) TELEFAX: (617) 542-8906
        (C) TELEX: 200154

- 114 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 1:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 2320
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:


```
GTTAACCACT CTTAATAATA ATGGGTTTTA TAGCGAAATA CACTTTTTTA TCGCGTGTTC     60

AATATTTGCG TTAGTTATTA TTTTTTTGGA ATGTAAATTC TCTCTAAACA CAGGTGATAT    120

TTATGTTGGA ATTGTGGTGT TGATTCTATT CTTATAATAT AACAAGAAAT GTTGTAACTG    180

ATAGATATAT TAAAAGATTA AATCGGAGGG GGAATAAAGC GTGCTAAGCA TCATCGTGAA    240

TATGATTACA GCGCCTGCGA TGGCATATAA CCGTATTGCG GATGGAGCGT CACGTGAGGA    300

CTGTGAAGCA CAATGCGATA TGTTCTGATT ATATGGCGAG TTTGCTTAAT GACATGTTTT    360

TAGCCGAACG GTGTCAAGTT TCTTAATGTG GTTGTGAGAT TTTCTCTTTA AATATCAAAA    420

TGTTGCATGG GTGATTTGTT GTTCTATAGT GGCTAAAGAC TTTATGGTTT CTGTTAAATA    480

TATATGCGTG AGAAAAATTA GCATTCAAAT CTATAAAAGT TAGATGACAT TGTAGAACCG    540

GTTACCTAAA TGAGCGATAG AGTGCTTCGG TAGTAAAAAT ATCTTTCAGG AAGTAAACAC    600

ATCAGGAGCG ATAGCGGTGA ATTATTCGTG GTTTTGTCGA TTCGGCATAG TGGCGATAAC    660

TGAATGCCGG ATCGGTACTG CAGGTGTTTA AACACACCGT AAATAATAAG TAGTATTAAG    720

GAGTTGTT                                                             728
```

```
ATG AAA AAT ATT ATT TTA TCC ACT TTA GTT ATT ACT ACA AGC GTT TTG     776
Met Lys Asn Ile Ile Leu Ser Thr Leu Val Ile Thr Thr Ser Val Leu
                5              10                  15

GTT GTA AAT GTT GCA CAG GCC GAT ACT AAC GCC TTT TCC GTG GGG TAT     824
Val Val Asn Val Ala Gln Ala Asp Thr Asn Ala Phe Ser Val Gly Tyr
            20              25                  30

GCA CGG TAT GCA CAA AGT AAA GTT CAG GAT TTC AAA AAT ATC CGA GGG     872
Ala Arg Tyr Ala Gln Ser Lys Val Gln Asp Phe Lys Asn Ile Arg Gly
        35              40                  45

GTA AAT GTG AAA TAC CGT TAT GAG GAT GAC TCT CCG GTA AGT TTT ATT     920
Val Asn Val Lys Tyr Arg Tyr Glu Asp Asp Ser Pro Val Ser Phe Ile
    50              55                  60

TCC TCG CTA AGT TAC TTA TAT GGA GAC AGA CAG GCT TCC GGG TCT GTT     968
Ser Ser Leu Ser Tyr Leu Tyr Gly Asp Arg Gln Ala Ser Gly Ser Val
65              70                  75                  80
```

```
GAG CCT GAA GGT ATT CAT TAC CAT GAC AAG TTT GAG GTG AAG TAC GGT        1016
Glu Pro Glu Gly Ile His Tyr His Asp Lys Phe Glu Val Lys Try Gly
                 85              90                   95

TCT TTA ATG GTT GGG CCA GCC TAT CGA TTG TCT GAC AAT TTT TCG TTA        1064
Ser Leu Met Val Gly Pro Ala Tyr Arg Leu Ser Asp Asn Phe Ser Leu
            100             105             110

TAC GCG CTG GCG GGT GTC GGC ACG GTA AAG GCG ACA TTT AAA GAA CAT        1112
Tyr Ala Leu Ala Gly Val Gly Thr Val Lys Ala Thr Phe Lys Glu His
        115             120             125

TCC ACT CAG GAT GGC GAT TCT TTT TCT AAC AAA ATT TCC TCA AGG AAA        1160
Ser Thr Gln Asp Gly Asp Ser Phe Ser Asn Lys Ile Ser Ser Arg Lys
    130             135             140

ACG GGA TTT GCC TGG GGC GCG GGT GTA CAG ATG AAT CCG CTG GAG AAT        1208
Thr Gly Phe Ala Trp Gly Ala Gly Val Gln Met Asn Pro Leu Glu Asn
145             150             155             160

ATC GTC GTC GAT GTT GGG TAT GAA GGA AGC AAC ATC TCC TCT ACA AAA        1256
Ile Val Val Asp Val Gly Tyr Glu Gly Ser Asn Ile Ser Ser Thr Lys
                165             170             175

ATA AAC GGC TTC AAC GTC GGG GTT GGA TAC CGT TTC TGA AAAGC             1300
Ile Asn Gly Phe Asn Val Gly Val Gly Tyr Arg Phe
            180             185
```

```
ATAAGCTATG CGGAAGGTTC GCCTTCCGCA CCGCCAGTCA ATAAAACAGG GCTTCTTTAC      1360

CAGTGACACG TACCTGCCTG TCTTTTCTCT CTTCGTCATA CTCTCTTCGT CATAGTGACG      1420

CTGTACATAA CATCTCACTA GCATAAGCAC AGATAAAGGA TTGTGGTAAG CAATCAAGGT      1480

TGCTCAGGTA GGTGATAAGC AGGAAGGAAA ATCTGGTGTA AATAACGCCA GATCTCACAA      1540

GATTCACTCT GAAAAATTTT CCTGGAATTA ATCACAATGT CATCAAGATT TTGTGACCGC      1600

CTTCGCATAT TGTACCTGCC GCTGAACGAC TACTGAAAAG TAGCAAGGTA TGTATTTTAT      1660

CCAGGAGAGC ACCTTTTTTG CGCCTGGCAG AAGTCCCCAG CCGCCACTAG CTCAGCTGGA      1720

TAGAGCATCA ACCTCCTAAG TTGATGGTGC GAGGTTCGAG GCCTCGGTGG CGGTCCAATG      1780

TGGTTATCGT ATAATGTTAT TACCTCAGTG TCAGGCTGAT GATGTGGGTT CGACTCCCAC      1840

TGACCACTTC AGTTTTGAAT AAGTATTGTC TCGCAACCCT GTTACAGAAT AATTTCATTT      1900

ATTACGTGAC AAGATAGTCA TTTATAAAAA ATGCACAAAA ATGTTATTGT CTTTTATTAC      1960

TTGTGAGTTG TAGATTTTTC TTATGCGGTG AATCCCCCTT TGCGGCGGGG CGTCCAGTCA      2020

AATAGTTAAT GTTCCTCGCG AACCATATTG ACTGTGGTAT GGTTCACCGG GAGGCACCCG      2080

GCACCGCAAT TTTTTATAAA ATGAAATTCA CACCCTATGG TTCAGAGCGG TGTCTTTTTA      2140

CATCAGGTGG GCAAGCATAA TGCAGGTTAA CTTGAAAGAT ACGATCAATA GCAGAAACCA      2200

GTGATTTCGT TTATGGCCTG GGGATTTAAC CGCGCCAGAG CGTATGCAAG ACCCTGGCGC      2260

GGTTGGCCGG TGATCGTTCA ATAGTGCGAA TATGAATGGT TACCAGCCGC CTGCGAATTC      2320
```

- 116 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 2:

  (i) SEQUENCE CHARACTERISTICS:

    (A) LENGTH: 53
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: single
    (D) TOPOLOGY: linear

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CATTTCTCAT TGATAATGAG AATCATTATT GACATAATTG TTATTATTTT ACG   53

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 3:

  (i) SEQUENCE CHARACTERISTICS:

    (A) LENGTH: 688
    (B) TYPE: nucleic acid
    (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear

  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAGCGCATTA TCAGATAAAT TGATTTATTT CTCACTTTCA TTCTATTTTC ATCAGGAATC 60

CCTGTGTCCT GTGCGGTAAT CTGCTGCTAT CGAGAACGAC AGACATCGCT AACAGTATAT 120

ATGGAAACAT CAAAAGAGAA GACGATAACA AGCCCAGGGC CATACATAGT TCGATTACTT 180

AACAGCTCAC TGAACGGCTG TGAGTTTCCA TTGCTGACAG GCCGAACACT CTTTGTGGTA 240

GGTCAGAGTG ATGCGCTCAC TGCTTCAGGT CAACTCCCTG ATATACCTGC CGATAGCTTT 300

TTTATCCCGC TGGACCATGG CGGAGTAAAT TTTGAAATCC AGGTGGATAC GGATGCGACC 360

GAAATTATAC TCCATGAGCT GAAAGAAGGA AATTCTGAAT CTCGTTCGGT GCAATTAAAT 420

ACGCCAATAC AGGTCGGTGA ATTGCTTATC CTGATTCGCC CGGAAAGCGA GCCGTGGGTG 480

CCCGAGCAGC CTGAGAAGTT AGAAACGTCT GCAAAAAAGA ACGAGCCGCG TTTTAAAAAC 540

GGAATTGTAG CAGCACTGGC CGGGTTTTTT ATATTGGGAA TTGGGACTGT GGGGACGTTA 600

TGGATACTTA ACTCGCCGCA GCGGCAGGCC CGAGAGCTCG ATTCGTTATT GGGGCAGGAG 660

AAGGAGCGTT TTCAGGTGTT GCCAGGCC   688

- 117 -

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 4:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 16
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATATCGCCC TGAGCA   16

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 5:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 4044
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGTTAACTCT TCGTTGAATA AAAAATGTCA ATGACGTTCC ATAATTCAGG AGATGAACTT   60

CACAAGTCAT TATATATAAC AGGAGGTGCT ATGAAACATC ATGCTTTTAT GCTTTGGTCA   120

TTACTTATTT TTTCATTCCA TGTTTTGGCC AGTTCAGGCC ATTGTTCTGG TTTACAACAG   180

GCATCATGGG ATATTTTTAT CTACGATTTT GGTAGTAAAA CCCCGCAACC ACCTACAAAT   240

ACTGATAAAA AGCAAGCCAG GCAGATTAGT TCACCGTCCT GCCCGACGAC AAAACCCATG   300

ATGTCCGCAC CAGTCAATGA CGCCAGGAAA GGGAATACTT TCTCCAGAAC ATAATGTTAT   360

TTATCTACAA TGGTGCCGAC GACTACTTTT AGCCACCCGG AAATCTTGAT TGCCATCAAA   420

TATAGCTGGC ATTATTTTTC CTGACGTGTA TAGTGCGCCT CGTTATCCCC ATTAAGGAAT   480

TTGTTTGTCT CGTAAAATGA CAGGAATTGT CAAAACCTTT GATTGTAAGA GCGGTAAAGG   540

TCTCATCACC CCCTCCGATG ACGCAAAGAT GTTCAGGTCC ACATTCAGC ATGTCGCCAA   600

CACGAAACAG AAGCGCTTAT CCCCGGTATA CGCGTTGAGT TTTATCGTAT TAATGGCCTC   660

CGCGGACCTA CCGCCGCCAA CGTTTATCTT TCATAATTCG TCACCCGGCA TTTTTCAGAA   720

AAATTTAGCG AGTACGTCTA CCTCCGCAGC CTGCTATGAG GCTTTGCCTG AAAGGCTGCA   780

GAATGTTTTC AGTGGCGAAA ATCTAAAAGA TTTATTTTGC TAATCAGTCC TGTGACCTCT   840

- 118 -

```
TTTATCATAT ATCGGGTGCC CCCCCTTCTC ACTTTGTTTA ACGTGAAGAA ATGTACAGCC      900

GTTTTTCACT GTGATAGCAT CTAATATTGC AAAAGTATTT AACGCTATAT ACCCATTGTC      960

ACAGGAGTGG CTGCGTGCGA GCTGAGCTAT TTAACCGAAG TATTTATGTG ATCATTGGAA     1020

TTATCTCTAT TGCCGCTCAA TGCTACGTCA TATTCAGTGG GTATAAATCG CCAATATAGT     1080

TGTAACGCTA TTTATTTTTA GGGTAATAAT TGAATGACTT TGCTTTCAGG AAAAACCACA     1140

CTGGTTCTCT GCCTCTCCTC TATTTTATGT GGATGTACGA CGAACGGCTT ACCCACACCT     1200

TATAGTATTA ATTTGTCGTT CCCGGTCATT ACACAAAACC AGATTAATTC CGGTGGTTAT     1260

TACATAAATG ACGCGGAACA AATTCGGACA ACTGATGGTC TGTGCCTTGA TGCAGGCCCA     1320

GATCAACAGA ATCGTTTGAC GCTGCGGGAG TGTAAGCATG TGCAATCTCA GCTTTTCTCA     1380

TTTCACCGAG ACAGAATCAC GCAGGGTGAG AAATGTCTGG ATGCCGCAGA CAAGGTACAA     1440

AAGAAGGCAC ACCAATCATT CTTTATTCAT GCACGGGTAA TGATAACCAG CGCTGGCTCA     1500

CTGATCATAA CAAAATTAAG GGGAAACAGA GCCGAAAATG CCTGGGCACA AATAGCATTA     1560

TTGTCAGAAA AGGCGACCCT GTTGTGTTGG CCGATTGCGA TTTTAGTCGC GCCCTGGAAT     1620

TTACCATCAG GTAGCAGGAC ACCGCTGTGA AGAGAGTGCC GCTAACCTCT TGACACGACA     1680

ACAGGTTAGC GACCTTTACT TCCACGTGCG ATCAATTTAC TTTACGTCCG CAACGTCAGG     1740

ATGACAAAAC GGCGGCTAAA CCTTGACACC AGTTATATAC CCAGCTTAAA TACTGGTCAT     1800

CCAACCAGTA AAAAGGAAAT GGCGATGTTC GTCGAACTCG TTTATGACAA GCGAAATGTT     1860

GAAGGTTTGC CAGGCGCACG CGAAATCATC CTCAATGAAC TCACAAAACG CGTACATCAA     1920

CTTTTTCCCG ATGCGCAAGT GAAAGTTAAG CCAATGCAGG CGAACGCATT AAACAGTGAC     1980

TGTACAAAAA CCGAGAAAGA ACGGCTGCAC CGTATGCTGG AAGAGATGTT TGAAGAGGCT     2040

GATATGTGGC TGGTCGCCGA ATAACGTCCC CTCCTGCGAA AGCCAACATG TCCGATCGAA     2100

AACAGCGCCC TGAGGCGCTG TCTGTGACGA TATAACGCAA ACGCTACCAC TCAGAACATG     2160

TTGTTGTTGA TACCTCAGAC CGGTATGTGG AACCGACATT CATCGCTTCA CTGGCCTGTC     2220

GGTATGAGTA GCCCTTATCA ACAATCAGCT GTGCGCATTC CAGCCTGAAA TCTGAAAGTA     2280

CGTTTGGTTT TGTTGTTTAT TAAGAGCCTA TCCCATTAGA CTCTTTTATT CGCCAAACTG     2340

GCTTTAACGA TTACGCCTAC TGGGATAGGT TCTAAACTTA TCATCAATAC GTAAAATACC     2400

TATTTACGAA CAAAAGTAA CAGGTAAAAA TCCGAAATAA AACCAGCATA ACTAAAACTT     2460

ACTGCAGATA TGCACACGCA TTATTACTAT GTTTCCAGGA TAGTCTCGAC CAGTCAAGAC     2520

TATCTATTTT ATATAAAAAG GGAAATACTT CACATGAATA AAATACATGT TACATATAAA     2580

AATCTCTTAC TTCCGATTAC CTTCATCGCG GCAACTCTAA TTAGCGCCTG TGATAACGAT     2640

AAAGATGCCA TGGCGGAAGC TGAAAAAAAT CAAGAGAAAT ACATGCAAAA AATCCAGCAA     2700

AAAGAGCACC AGCAATCAAT GTTCTTTTAC GACAAAGCCG AAATGCAAAA AGCTATTGCC     2760
```

- 119 -

```
AATATCAACG CAAAAGGTGG AGCCAATCTT GCGATTATTG AAGTCCGTTT CTTCAAGGGC    2820
GGGTATTCAT TCATTCGACA AAGCGTTAAC ACCCCTGCTA AAGTAGAGGT GTTTAAATTT    2880
AACAACGGCT ACTGGGGGGG ACCTTCGCCT GTCAATTTAA CCATCTTTGG CACTATAACA    2940
GAGGAGCAAA AACAAGAAGC ACTAAAAGAG GCTTTATTCA AATTCGACTC GATCAATTTC    3000
AGCATTATAC CAGAGCGTAT TCAGGAAACA ATTAAACGCG CTAACGCCAG TGGCATCATT    3060
TCCGTTACGG AAGATAGCGA TATCGTTGTA CGAGCAGAGA TAGCTCATAA TGGCGAATTC    3120
GTCTATGACA TTACCATCAC TGCTAAAAAT ACAGCACGTG CGGTAATGAC CTTAAATAAG    3180
GATGGTTCTA TTGCCGGATA TGAGATCAAA GAACCTTTCG CCCCAAAAAA AGAAGCCGAA    3240
AAAGCACAGC AACTTGTTGA ACAATCGAGA AAAGACATTG AAAGTCCAGC GTAAAAAAGC    3300
AGCTGGAAAG ATGAACGAAA TACAGCAGAC ATTTAAAAAT AGCAGGCGAT ACAAACATTG    3360
ATAAAAATTA TAGCGCGAAA GAGCGCGTGC CAGGTACTAA GGCACTGCTT GAAGACAGCG    3420
AATCGCTATT TCATTCTCTG ACACTGTAAT TTTTCGTACT CAAGATGTTT ATTTATTGAG    3480
TCTTTTGTGG ATAACCAGGT GAAGTTATGT GACGCCAGGA ATCTATTCCA GCGGGCGTAC    3540
TTGTTGGAGC CAGTGTGAAG CCGGGCAGCG CGCAGAAACC GGAGCGTATA CGTTGTACGT    3600
AAGAATTTCG AGCACTGCCC GACCTAAAAA TGATGAATAA AATAGATATT TTAAAGAGGT    3660
AATATGAAGA ATTTTTTCAA AATAATTACT GATTTCATCG CGGATATTTC CCTTGATCTA    3720
TTTGCTATAT TTTTATGCAT GTTATTCGTA TACAAAACAG GACCATCAAT TGGTGTGATA    3780
TCATTTTTTA TTGCATTAAT TATTTATATC ATTCTTCATT TTTTTTTACT CATTTCTTGA    3840
AAAAATCATA AAAAAAATAT TCAAATAAGT ATTTAAAATT ATTGTTTTGT GGTACAAATT    3900
CAGCGCAATA AAACAGAGCA ACTAAAAAAA ATTAGGCGTA GCGAAGTGGA AAAGGACTGT    3960
CATGTACTGG ACCGTGAGCT GGTCGGGAGA GCAATGTACG GGAAAGAGCG AAATACTGTC    4020
ATTGATATGA GCAGGAATAT CGAT                                          4044
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 6:

     (i) SEQUENCE CHARACTERISTICS:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 87 amino acids
          (B) TYPE: amino acid
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: peptide

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

- 120 -

```
Met Lys His His Ala Phe Met Leu Trp Ser Leu Leu Ile Phe Ser Phe
1               5               10              15
His Val Leu Ala Ser Ser Gly His Cys Ser Gly Leu Gln Gln Ala Ser
            20              25              30
Trp Asp Ile Phe Ile Tyr Asp Phe Gly Ser Lys Thr Pro Gln Pro Pro
        35              40              45
Thr Asn Thr Asp Lys Lys Gln Ala Arg Gln Ile Ser Ser Pro Ser Cys
    50              55              60
Pro Thr Thr Lys Pro Met Met Ser Ala Pro Val Asn Asp Ala Arg Lys
65              70              75              80
Gly Asn Thr Phe Ser Arg Thr
            85
```

(2) **INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER:**   7:

   (i) SEQUENCE CHARACTERISTICS:

      (A) LENGTH: 178 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

   (ii) MOLECULE TYPE: peptide

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7

```
Met Thr Leu Leu Ser Gly Lys Thr Thr Leu Val Leu Cys Leu Ser Ser
1               5               10              15
Ile Leu Cys Gly Cys Thr Thr Asn Gly Leu Pro Thr Pro Tyr Ser Ile
            20              25              30
Asn Leu Ser Phe Pro Val Ile Thr Gln Asn Gln Ile Asn Ser Gly Gly
        35              40              45
Tyr Tyr Ile Asn Asp Ala Glu Gln Ile Arg Thr Thr Asp Gly Leu Cys
    50              55              60
Leu Asp Ala Gly Pro Asp Gln Gln Asn Arg Leu Thr Leu Arg Glu Cys
65              70              75              80
Lys His Val Gln Ser Gln Leu Phe Ser Phe His Arg Asp Arg Ile Thr
            85              90              95
Gln Gly Glu Lys Cys Leu Asp Ala Ala Asp Lys Val Gln Lys Lys Ala
            100             105             110
His Gln Ser Phe Phe Ile His Ala Arg Val Met Ile Thr Ser Ala Gly
        115             120             125
Ser Leu Ile Ile Thr Lys Leu Arg Gly Asn Arg Ala Glu Asn Ala Trp
    130             135             140
Ala Gln Ile Ala Leu Leu Ser Glu Lys Ala Thr Leu Leu Cys Trp Pro
145             150             155             160
```

- 121 -

```
Ile Ala Ile Leu Val Ala Pro Trp Asn Leu Pro Ser Gly Ser Arg Thr
              165                 170                 175
Pro Leu
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 8:

   (i) SEQUENCE CHARACTERISTICS:

      (A) LENGTH: 79 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8

```
Met Phe Val Glu Leu Val Tyr Asp Lys Arg Asn Val Glu Gly Leu Pro
1             5             10                 15
Gly Ala Arg Glu Ile Ile Leu Asn Glu Leu Thr Lys Arg Val His Gln
            20             25                 30
Leu Phe Pro Asp Ala Gln Val Lys Val Lys Pro Met Gln Ala Asn Ala
        35             40             45
Leu Asn Ser Asp Cys Thr Lys Thr Glu Lys Glu Arg Leu His Arg Met
    50             55             60
Leu Glu Glu Met Phe Glu Glu Ala Asp Met Trp Leu Val Ala Glu
65             70             75
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 9:

   (i) SEQUENCE CHARACTERISTICS:

      (A) LENGTH: 246 amino acids
      (B) TYPE: amino acid
      (D) TOPOLOGY: linear

  (ii) MOLECULE TYPE: peptide

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9

```
Met Asn Lys Ile His Val Thr Tyr Lys Asn Leu Leu Leu Pro Ile Thr
1             5             10                 15
Phe Ile Ala Ala Thr Leu Ile Ser Ala Cys Asp Asn Asp Lys Asp Ala
            20             25                 30
Met Ala Glu Ala Glu Lys Asn Gln Glu Lys Tyr Met Gln Lys Ile Gln
            35             40                 45
```

- 122 -

```
Gln Lys Glu His Gln Gln Ser Met Phe Phe Tyr Asp Lys Ala Glu Met
    50                  55                  60

Gln Lys Ala Ile Ala Asn Ile Asn Ala Lys Gly Gly Ala Asn Leu Ala
65              70                  75                      80

Ile Ile Glu Val Arg Phe Phe Lys Gly Gly Tyr Ser Phe Ile Arg Gln
                85                  90                  95

Ser Val Asn Thr Pro Ala Lys Val Glu Val Phe Lys Phe Asn Asn Gly
                100             105             110

Tyr Trp Gly Gly Pro Ser Pro Val Asn Leu Thr Ile Phe Gly Thr Ile
            115             120             125

Thr Glu Glu Gln Lys Gln Glu Ala Leu Lys Glu Ala Leu Phe Lys Phe
    130             135             140

Asp Ser Ile Asn Phe Ser Ile Ile Pro Glu Arg Ile Gln Glu Thr Ile
145             150             155             160

Lys Arg Ala Asn Ala Ser Gly Ile Ile Ser Val Thr Glu Asp Ser Asp
                165             170             175

Ile Val Val Arg Ala Glu Ile Ala His Asn Gly Glu Phe Val Tyr Asp
            180             185             190

Ile Thr Ile Thr Ala Lys Asn Thr Ala Arg Ala Val Met Thr Leu Asn
        195             200             205

Lys Asp Gly Ser Ile Ala Gly Tyr Glu Ile Lys Glu Pro Phe Ala Pro
    210             215             220

Lys Lys Glu Ala Glu Lys Ala Gln Gln Leu Val Glu Gln Ser Arg Lys
225             230             235             240

Asp Ile Glu Ser Pro Ala
            245
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 10:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 3700
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10

```
TTTTGGTTTG CTGCCGTTTG GGATAACTGC ATAGAGAGCG GCCAAGTCGC TTGCGGTCGG      60
TATCTCGAGT ATATCGAAAT CCATGTGGCC ATTGACCTCT TCAAGCGCTC ACGTTAACTA     120
CCTGCTCTTT TTTGAGCACC AACATCCCAG GTTCGTCACA GTAAATCGTA TCGTGATTAT     180
```

```
TGCTAATCGT CAGTTTACCG CTCCGAAAGC AAACTAAAGT GAAACTGCTT ACATAAAGAT        240

TTTTGATGGT AACCTGCTGA GTCTGACTTT TAATTTGCTG CCGGGTATTT GTCAAAAGTG        300

ATTTTAATTT CTGTAAGTTA TCTGCGGCAG GACGCTGATG ACTATTACTT ACAAAGGTTA        360

CATTTTCCAT ATTATCCCTT TGTTGAACTT ATTTTAATGT TCCTTACTGG TATCCTACTG        420

AAAAAATCTG AGTTGTAAAT GCTCTTTATT AGCGTGTGTT GGCAATGGTC TGATTGTTAC        480

ACCAAAAGAA CCCAAATTTG GGTAATTTAT CTACAGTAGT TTAAGCCCCA ATGGGGATGA        540

TGGTTCTTTT AATATGTGTT GAGACGCATT ATACAGAATA AATTGATTTT ATTTCTCACT        600

TTTCATTCTA TTTTCATCAG GAATCCCTGT GTCCTGTGCG GTAATCTGCT GCTATCGAGG        660

AACGACAGAC ATCGCTAACA GTATATATGG AAACATCAAA AGAGAAGACG ATAACAAGCT        720

TTCCAGGGCC ATACATAGTT CGATTACTTA ACAGCTCACT GAACGGCTGT GAGTTTCCAT        780

TGGGCCTGAC AGGCCGAACA CTCTTTGTGG TAGGTCAGAG TGATGCGCTC ACTGCTTCAG        840

GTCAATGTGA TAGCTCCCTG ATATACCTGC CGATAGCTTT TTTATCCCGC TGGACCATGG        900

CGGAGTAAAT TTTAGGGAAA TCCAGGTGGA TACGGATGCG ACCGAAATTA TACTCCATGA        960

GCTGAAAGAA GGAAATTATG TCTGAATCTC GTTCGGTGCA ATTAAATACG CCAATACAGG       1020

TCGGTGAATT GCTTATCCTG TGATTCGCCC GGAAAGCGAG CCGTGGGTGC CCGAGCAGCC       1080

TGAGAAGTTA GAAACGTCTG CATAAAAAAG AACGAGCCGC GTTTTAAAAA CGGAATTGTA       1140

GCAGCACTGG CCGGGTTTTT TATAGAAAGT TGGGAATTGG GACTGTGGGG ACGTTATGGA       1200

TACTTAACTC GCCGCAGCGG CAGGCCGCAG GTGTAAGAGC TCGATTCGTT ATTGGGGCAG       1260

GAGAAGGAGC GTTTTCAGGT GTTGCCAGGC CGGGACGGAA AATGCTCTAT GTCGCTGCGC       1320

AAAATGAAAG AGATACGTTG TGGGCTCGTC AGGTTTTAAA TAGCGAGGGG CGATTATGAT       1380

AAAAATGCGC GAGTGATTAA CGAAAACGAA GAAAATAAGC GTAGAATCTC TATCTGGCTG       1440

GATACCTATT ATCCGCAGCT GGCTTATTAT CGGATTCATT TCGATTAGAG CCGCGTAAAC       1500

CCGTTTTCTG GCTAAGCCGC CAGCGAAACA CGATGAGCAA GAAAGAGTCT CGAGGTGTTA       1560

AGTCAAAAGC TGAGAGCGCT AATGCCTTAC GCGGATTCGG TTAACATCAA ACGTTGATGG       1620

ACGATGTTAC CGCAGCAGGC CAGGCGGAAG CGGGGCTAAA ACAGCAGGCG TTAAGAAGAT       1680

TACCTTATTC CCGCAGGAAT CATAAGGGGG GCGTAACGTT TGTTATTCAG GGGGCGCTCG       1740

GTGAGATGAT GTAGAAATAC TCAGAGCCCG TCAATTTGTC GATAGCTATT ACCGCACATG       1800

GGGAATGGGA CGCTATGTGC AGTTTGCGAT CGAATTAAAA GATGACTGGC TCAAGGGGCG       1860

CTCATTTGAG CAGTACGGGG CGGAAGGTTA TATCAAAATG AGCCCAGGCC ATTGGTATTT       1920

CCCAAGCCCA GAGGGCTTTA ATTTAACGTA AATAAGGAAG TCATTATGGC AACACCTTGG       1980

TCAGGCTATC TGGATATGGA CGTCTCAGCA AAATTTGATA CGGGCGTTGA TAATCTACAA       2040

ACGCAGGTAA CAGAGGCGAT GTTACTGGAT AAATTAGCAG CAAAACCCTC CGATCCGGCG       2100
```

- 124 -

```
CTACTGGCGG CGTATCAGAG TAAGAAAAAC TCTCGGAATA TAACTTGTAC CGTAACGCGC   2160

AATCGAACAC GGTAAAAGTC TTTAAGGATA TGATTGATGC TGCCATTATT CAGAACTTCC   2220

GTTAATCAGT TATAAGGTGG ATTATGTCGA TTAAGCAACT ATTGTCCCTG AGAATGCCGT   2280

TATAGGGCAG GCGGTCAATA TCAGGTCTAT GGAAATAGAA CGGACATTGT CTCGCTGGAT   2340

GACCGGCTAC TCCAGGCTTT TTCTGGTTCG GCGATTGCCT AGAAACGGCT GTGGATAAAC   2400

AGACGATTAC CAACAGGATT GAGGACCCTA ATCTGGTGAC GGATTATTTC CTAAAGAGCT   2460

GGCTATTTCG CAAGAGATGA TTTCAGATTA TAACCTGTAT GTTTCTATGA GGTCAGTACC   2520

CTTACTCGTA AAGGAGTCGG GGCTGTTGAA ACGCTATTAC GCTCATGATT CTTGGATGTC   2580

GATATCTATA TACTTTTCTG CTGGTAATGA CCCTTGCCGG CTGTAAGGAT AAGGATCTTA   2640

GCTTTTAAAA GGACTGGACC AGGAACAGGC TAATGAGGTC ATTGCCGTTC TGCAAATGCA   2700

CAGAAATATA GAGGCGAATA AAATTGATAG CGGAAAATTG GGCTATAGCA TTACCGTTGC   2760

TGAGCAGGTA CTGATTTTAC CGCTGCGGTG TACTGGATTA AAACTTATCA GCTTCCTCCC   2820

CGGCCACGGG TAATTGGAAA TAGCGCAGAT GTTCCCGGCG GATTCGCTGG TATCGTCTCC   2880

GCGAGCTGAA AAGGAAAACC AGGTTATATT CGGCTATTGA ACAGCGACTG GAACAGTCAT   2940

TACAGACGAT GGAGGGCGAT GTGCTCTCCG CCAGGGTCCA TATTAGTTAT GATATTGATG   3000

CTGGTGAAAA TGGCCGCCCG CAAGGCAAAA CCTGTTCATC TGTCGGCATT AGCCGTATAT   3060

GAACGAGGTT CGCCGCTTGC GCATCAAGAA GATCAGCGAT ATCAAGCGTT TCTTAAAGAA   3120

TAGTTTTGCC GATGTGGATT ATGACAACAA TTTCTGTTGT GTTGTCAGAA CGTTCTGATG   3180

CCCAATTACA GGCTCCCGGC ACACCAGTAA AAGTAACGTA ATTCTTTTGC AACCAGTTGG   3240

ATTGTTTTGA TTATTTTGTT ATCCGTGATG TCAGATACAG GCTTTGGCGT CTGGTATTAC   3300

AAAAACCATT ATGCCCGCAA TAAGAAAGGC ATAACGGGGA GTACTGATGA TAAGGCGAAA   3360

TCGTCAAATG AATAGGCAGC CATTACCCAT TATCTGGCAA AGAATCATTT TTGATCCGTT   3420

ATCGTATATC CATCCTCAGC GGTTGCAGAT AGCGCCGGAA ATGATTGTCA GACCGCGCCA   3480

CGCGAAATGA GTTAATACTG GCGGCATGGC GGCGGCTTAA GAACGGAGAA AAGGAGTGTA   3540

TTCAAAACTC ACTGACGCAG CTGTGGCTGC TCAGTGGCGC CGACTGCCGC AAGTAGCGTA   3600

TTTACTAAAC TGAGAGCCGA TCTGGCAAGG CAGGGAGCCT TGCTTGGCCT AGCCGGATTG   3660

GGCGAAATGA GTTAATACTG GCGGCATGGC GGCTTGCCAT                         3700
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 11:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 392 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11

```
Met Glu Thr Ser Lys Glu Lys Thr Ile Thr Ser Pro Gly Pro Tyr Ile
 1               5                  10                      15

Val Arg Leu Leu Asn Ser Ser Leu Asn Gly Cys Glu Phe Pro Leu Leu
            20                  25                  30

Thr Gly Arg Thr Leu Phe Val Val Gly Gln Ser Asp Ala Leu Thr Ala
        35              40              45

Ser Gly Gln Leu Pro Asp Ile Pro Ala Asp Ser Phe Phe Ile Pro Leu
    50              55              60

Asp His Gly Gly Val Asn Phe Glu Ile Gln Val Asp Thr Asp Ala Thr
65              70              75              80

Glu Ile Ile Leu His Glu Leu Lys Glu Gly Asn Ser Glu Ser Arg Ser
                85              90              95

Val Gln Leu Asn Thr Pro Ile Gln Val Gly Glu Leu Leu Ile Leu Ile
            100             105             110

Arg Pro Glu Ser Glu Pro Trp Val Pro Glu Gln Pro Glu Lys Leu Glu
        115             120             125

Thr Ser Ala Lys Lys Asn Glu Pro Arg Phe Lys Asn Gly Ile Val Ala
    130             135             140

Ala Leu Ala Gly Phe Phe Ile Leu Gly Ile Gly Thr Val Gly Thr Leu
145             150             155             160

Trp Ile Leu Asn Ser Pro Gln Arg Gln Ala Ala Glu Leu Asp Ser Leu
            165             170             175

Leu Gly Gln Glu Lys Glu Arg Phe Gln Val Leu Pro Gly Arg Asp Lys
            180             185             190

Met Leu Tyr Val Ala Ala Gln Asn Glu Arg Asp Thr Leu Trp Ala Arg
        195             200             205

Gln Val Leu Ala Arg Gly Asp Tyr Asp Lys Asn Ala Arg Val Ile Asn
    210             215             220

Glu Asn Glu Glu Asn Lys Arg Ile Ser Ile Trp Leu Asp Thr Tyr Tyr
225             230             235             240

Pro Gln Leu Ala Tyr Tyr Arg Ile His Phe Asp Glu Pro Arg Lys Pro
            245             250             255

Val Phe Trp Leu Ser Arg Gln Arg Asn Thr Met Ser Lys Lys Glu Leu
            260             265             270

Glu Val Leu Ser Gln Lys Leu Arg Ala Leu Met Pro Tyr Ala Asp Ser
        275             280             285

Val Asn Ile Thr Leu Met Asp Asp Val Thr Ala Ala Gly Gln Ala Glu
        290             295             300
```

- 126 -

```
Ala Gly Leu Lys Gln Gln Ala Leu Pro Tyr Ser Arg Arg Asn His Lys
305             310             315                 320

Gly Gly Val Thr Phe Val Ile Gln Gly Ala Leu Asp Asp Val Glu Ile
            325             330                 335

Leu Arg Ala Arg Gln Phe Val Asp Ser Tyr Tyr Arg Thr Trp Gly Gly
            340             345             350

Arg Tyr Val Gln Phe Ala Ile Glu Leu Lys Asp Asp Trp Leu Lys Gly
    355             360             365

Arg Ser Phe Gln Tyr Gly Ala Glu Gly Tyr Ile Lys Met Ser Pro Gly
    370             375             380

His Trp Tyr Phe Pro Ser Pro Leu
385             390
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 12:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 80 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12

```
Met Ala Thr Pro Trp Ser Gly Tyr Leu Asp Asp Val Ser Ala Lys Phe
 1               5               10              15

Asp Thr Gly Val Asp Asn Leu Gln Thr Gln Val Thr Glu Ala Leu Asp
            20              25              30

Lys Leu Ala Ala Lys Pro Ser Asp Pro Ala Leu Leu Ala Ala Tyr Gln
        35              40              45

Ser Lys Leu Ser Glu Tyr Asn Leu Tyr Arg Asn Ala Gln Ser Asn Thr
    50              55              60

Val Lys Val Phe Lys Asp Ile Asp Ala Ala Ile Ile Gln Asn Phe Arg
65              70              75              80
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 13:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 101 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- 127 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13

```
Met Ser Ile Ala Thr Ile Val Pro Glu Asn Ala Val Ile Gly Gln Ala
 1               5               10                  15

Val Asn Ile Arg Ser Met Glu Thr Asp Ile Val Ser Leu Asp Asp Arg
            20              25                  30

Leu Leu Gln Ala Phe Ser Gly Ser Ala Ile Ala Thr Ala Val Asp Lys
        35              40              45

Gln Thr Ile Thr Asn Arg Ile Glu Asp Pro Asn Leu Val Thr Asp Pro
    50              55              60

Lys Glu Leu Ala Ile Ser Gln Glu Met Ile Ser Asp Tyr Asn Leu Tyr
65              70              75              80

Val Ser Met Val Ser Thr Leu Thr Arg Lys Gly Val Gly Ala Val Glu
            85              90              95

Thr Leu Leu Arg Ser
        100
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 14:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 252 amino acids
        (B) TYPE: amino acid
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14

```
Met Ile Arg Arg Tyr Leu Tyr Thr Phe Leu Leu Val Met Thr Leu Ala
 1               5               10                  15

Gly Cys Lys Asp Lys Asp Leu Leu Lys Gly Leu Asp Gln Glu Gln Ala
            20              25                  30

Asn Glu Val Ile Ala Val Leu Gln Met His Asn Ile Glu Ala Asn Lys
        35              40              45

Ile Asp Ser Gly Lys Leu Gly Tyr Ser Ile Thr Val Ala Glu Pro Asp
    50              55              60

Phe Thr Ala Ala Val Tyr Trp Ile Lys Thr Tyr Gln Leu Pro Pro Arg
65              70              75              80

Pro Arg Val Glu Ile Ala Gln Met Phe Pro Ala Asp Ser Leu Val Ser
            85              90              95

Ser Pro Arg Ala Glu Lys Ala Arg Leu Tyr Ser Ala Ile Glu Gln Arg
        100             105             110
```

- 128 -

```
Leu Glu Gln Ser Leu Gln Thr Met Glu Gly Val Leu Ser Ala Arg Val
        115             120             125
His Ile Ser Tyr Asp Ile Asp Ala Gly Glu Asn Gly Arg Pro Pro Lys
    130             135             140
Pro Val His Leu Ser Ala Leu Ala Val Tyr Glu Arg Gly Ser Pro Leu
145             150             155             160
Ala His Gln Ile Ser Asp Ile Lys Arg Phe Leu Lys Asn Ser Phe Ala
                165             170             175
Asp Val Asp Tyr Asp Asn Ile Ser Val Val Leu Ser Glu Arg Ser Asp
            180             185             190
Ala Gln Leu Gln Ala Pro Gly Thr Pro Val Lys Arg Asn Ser Phe Ala
        195             200             205
Thr Ser Trp Ile Val Leu Ile Ile Leu Leu Ser Val Met Ser Ala Gly
    210             215             220
Phe Gly Val Trp Tyr Tyr Lys Asn His Tyr Ala Arg Asn Lys Lys Gly
225             230             235             240
Ile Thr Ala Asp Asp Lys Ala Lys Ser Ser Asn Glu
            245             250
```

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 15:

    (i) SEQUENCE CHARACTERISTICS:

        (A) LENGTH: 818
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: double
        (D) TOPOLOGY: linear

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
CATAACAACT CCTTAATACT ACTTATTATT TACGGTGTGT TTAAACACCT GCAGTACCGA    60
TCCGGCATTC AGTTATCGCC ACTATGCCGA ATCGACAAAA CCACGAATAA TTCACCGCTA   120
TCGCTCCTGA TGTGTTTACT TCCTGAAAGA TATTTTTACT ACCGAAGCAC TCTATCGCTC   180
ATTTAGGTAA CCGGTTCTAC AATGTCATCT AACTTTTATA GATTTGAATG CTAATTTTTC   240
TCACGCATAT ATATTTAACA GAAACCATAA AGTGTTTAGC CACTATAGAA CAACAAATCA   300
CCCATGCAAC ATTTTGATAT TTAAAGAGAA AATCTCACAA CCACATTAAG AAACTTGACA   360
CCGTTCGGCT AAAAAACATG TCATTAAGCA AACTCGCCAT ATAATCAGAA CATATCGCAT   420
TGTGCTTCAC AGTCCTCACG TGACGCTCCA TCCGCAATAC GGTTATATGC CATCGCAGGC   480
GCTGTAATCA TATTCACGAT GATGCTTAGC ACGCTTTATT CCCGCTCCGA TTTAATCTTT   540
TAATATATCT ATCAGTTACA ACATTTCTTG TTATATTATA AGAATAGAAT CAACACCACA   600
ATTCCAACAT AAATATCACC TGTGTTTAGA GAGAATTTAC ATTCCAAAAA AATAATAACT   660
```

- 129 -

```
AACGCAAATA TTGAACACGC GATAAAAAAG TCTATTTCGC TATAAAACCC ATTATTATTA     720

AGAGTGGTTA ACTCTTCGTT GAATAAAAAA TGTCAATGAC GTTCCATAAT TCAGGAGATG     780

AACTTCACAA GTCATTATAT ATAACAGGAG GTGCTATG                             818
```

TABLE 12

Bacterial strains.

Strain	Genotype	Source
S. typhimurium		
14082s	Wild type	ATCC
CS019	*phoN2 zxx*::6251Tn*10*d-Cm	25
CS015	*phoP-102*::Tn*10*d-Cm	25
AD154	*phoP12 purB1744*::Tn10	3
TT13208	*phoP105*::Tn*10*d	26
CS585	*pagD1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS586	*pagD1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS619	*pagE1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS620	*pagE1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1599	*pagF1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1600	*pagF1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS334	*pagG1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS335	*pagG1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1488	*pagH1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1489	*pagH1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS2054	*pagI1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS2055	*pagI1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1074	*pagJ1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1075	*pagJ1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS767	*pagK1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS768	*pagK1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS993	*pagL1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS994	*pagL1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1845	*pagM1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1846	*pagM1*:Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS728	*pagN1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS729	*pagN1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1194	*pagO1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1195	*pagO1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1247	*pagP1*::Tn*phoA phoN2 zxx*::6215Tn*10*d-Cm	This study
CS1248	*pagP1*::Tn*phoA phoP105*::Tn*10*d *phoN2 zxx*::6215Tn*10*d-Cm	This study
AK3011-3314	Collection of Randomly spaced Tn*10* Δ16Δ17 insertions	18
E. coli		
SM10(pRT291)	Contains plasmid pRT291 (Tn*phoA*) derived from pRK290 selecting for Tet^r and Km^r	37
MM294(pPH1JI)	Contains Gm^r plasmid pPH1JI, which is in incompatible with pRK290	37

[3] Behlau et al., 1993, J. Bacteriol., 175:4475-84

[18] Lehrer et al., 1991, Cell, 64:229-30

[25] Miller et al., 1989, Proc. Natl. Acad. Sci. USA, 86:5054-58

[26] Miller et al., 1990, J. Bacteriol., 172:2485-90

[37] Taylor et al., 1989, J. Bacteriol., 171:1870-78

-131-

TABLE 13

Comparison of *pag::phoA* activity in strains with wild type and null *phoP⁻* loci.

| Allele | Activity (Units of AP)[a] | | | | |
| | Logarithmic growth | | Stationary growth | | Fold Reduction[b] |
	PhoP+	PhoP-	PhoP+	PhoP-	
pagD1::Tn*phoA*	32	2	79	9	16
pagE1::Tn*phoA*	96	2	108	3	48
pagF1::Tn*phoA*	89	4	276	10	22
pagG1::Tn*phoA*	35	1	65	6	35
pagH1::Tn*phoA*	35	5	38	6	7
pagI1::Tn*phoA*	12	2	24	8	6
pagJ1::Tn*phoA*	123	8	944	88	15
pagK1::Tn*phoA*	30	3	123	26	10
pagL1::Tn*phoA*	7	1	35	4	7
pagM1::Tn*phoA*	92	11	439	130	8
pagN1::Tn*phoA*	23	1	58	2	23
pagO1::Tn*phoA*	31	2	54	12	16
pagP1::Tn*phoA*	38	1	27	3	38

[a] The AP activity values are presented in units as defined by Miller for ß-galactosidase (24). The values are representative of experiments (performed in duplicate) that were repeated on three separate occasions. PhoP⁺ denotes the *pag*::Tn*phoA* insertion in strain CS019 containing a wild type *phoP* locus. PhoP⁻ denotes an isogenic strains carrying the *phoP105*::Tn*10* allele.

[b] Values of fold reduction in enzymatic activity represent the decrease in AP activity on acquisition of the null *phoP105* allele. These were calculated from logarithmic growth phase cultures and rounded to the nearest whole number.

TABLE 14

The effects of *pag:: phoA* gene fusions on *Salmonella* mouse virulence.

Strain	Genotype	LD_{50}[a]	MSI[b]	Reference
14028s	Wild type	< 20	6.13	25
CS015	*phoP102::Tn10-Cam*	7.0×10^5	0.40	25
CS585	*pagD1::TnphoA*	4.0×10^5	0.01	15
CS1074	*pagJ1::TnphoA*	4.0×10^3	0.56	This study
CS767	*pagK1::TnphoA*	9.0×10^4	0.04	This study
CS1845	*pagM1::TnphoA*	3.0×10^4	0.09	This study

[a] The 50% lethal dose was determined by intraperitoneal injection of ten mice per dilution using the method of Reed and Muench (31).

[b] The Macrophage Survival Index (MSI) was determined by dividing the mean *Salmonella* CFU recovered from macrophage cultures (performed in triplicate) 24 hours after the addition of gentamicin by the mean *Salmonella* CFU recovered from macrophages 1 hour after gentamicin was added.

16   Kier et al., 1979, J. Bacteriol., 138:155-61

25   Miller et al., 1989, Proc. Natl. Acad. Sci. USA, 86:5054-58

TABLE 15

Plasmids, strains and relevent properties

S. typhimurium strains	Relevent genotypes/information	MSI[a]	Source[b]
ATCC14028	Wild type	3.90	ATCC
CS019	phoN2 zxx::6251Tn10d-Cm		(31)
CS585	CS019, pagD::TnphoA	0.002	(4)
CS586	CS585; phoP105::Tn10d-Tet		(4)
JSG205	ATCC14028, msgA::MudJ	0.01	This work
JSG225	JSG205, phoP105::Tn10d-Tet		This work
CS811	CS019, envE::TnphoA		This work
CS812	CS811, phoP105::Tn10d-Tet		This work
CS100	ATCC14028, phoP105::Tn10d-Tet	0.01	derivitive of TT13208
JSG232	JSG205, envF::pGPP2		This work
JSG234	CS019, envF::pGPP2		This work
JSG235	JSG234, phoP105::Tn10d-Tet		This work
JSG244	JSG205, phoP105::Tn10d-Tet		This work
CS099	ATCC14028;zxx3024::Tn10Δ16Δ17pol-2(Whitfield polA amber)		This work
Other salmonellae			
Ty2	Vi positive		FDA
Salmonella paratyphi A	ATCC 9150		ATCC
Salmonella paratyphi C	ATCC 13428		ATCC
Salmonella enteriditis	Clinical isolate		VRI
E. coli Strains			
SM10λpir	thi-1 thr-1 leuB6 supE44 tonA21 lacY1recA::RP4-2-Tc::Mu		
DH5α	F- Ø 80dlacZΔM15 Δ(lacZYA-argF)U169endA1recA1hsdR17deoRthi-1supE44λ-gyrA96relA1		
Other Enterobacteriaceae			
Yersinia enterocolitica	Clinical isolate		MGH bacteriology lab
Vibrio cholerae	Clinical isolate		Peruvian epidemic

Campylobacter fetus	Clinical isolate	MGH bacteriology lab
Citrobater freundii	Clinical isolate	MGH bacteriology lab
Klebsiella pneumoniae	Clinical isolate	MGH bacteriology lab
Shigella flexneri	Clinical isolate	MGH bacteriology lab
Shigella sonnei	Clinical isolate	MGH bacteriology lab
Morganella morganii	Clinical isolate	MGH bacteriology lab
Providencia stuartii	Clinical isolate	MGH bacteriology lab

Plasmids

pWPL17	pBR322 containing a 2.8 Kb *Hpa*I fragment from pWP061	This wor
pCAA9	pWPL17 containing a Tn*phoA* insertion in *envF*	This wor
pGP704	*pir*-dependent suicide vector	(34)
pGPP2	pGP704 containing the cloned *envF::phoA* gene fusion	This wor
pWP061	Cosmid clone containing the *pagC* region	(36)

a MSI (macrophage survival index) is calculated by dividing the number of surviving organisms at 2
hours post-infection by the number of cell asscociated organisms present after the 30 minute infectior

b MGH, Massachusetts General Hospital, ATCC, American Type Culture Collection, FDA, Food ar
Drug Administration; VRI, Virus Research Institute

4   Belden et al., 1989, Infect. Immun., 57:1-7
31   Miller et al., 1989, Proc. Natl. Acad. Sci. USA, 86:5054-58
34   Miller et al., 1988, J. Bacteriol., 170:2575-83
36   Pulkkinen et al., 1991, J. Bacteriol., 173:86-93

TABLE 16

Alkaline phosphatase and ß-galactosidase gene fusion activity

Strain	Relevent Genotype	gene fusion activity[a]
JSG205	*msgA*:MudJ	461(B)
JSG244	*phoP*105::Tn*10*d-Tet *msgA*:MudJ	415(B)
JSG226	*envE*::Tn*phoA*	50(A)
JSG229	*phoP*105::Tn*10*d-Tet *envE*::Tn*phoA*	60(A)
JSG204	*pagD*::Tn*phoA*	76(A)
JSG225	*phoP*105::Tn*10*d-Tet *pagD*::Tn*phoA*	9(A)
JSG234	*envF*::pGPP2	16(A)
JSG235	*phoP*105::Tn*10*d-Tet *envF*::pGPP2	19(A)
JSG232	*msgA*::MudJ *envF*::pGPP2	10(A)

[a] (A) AP (alkaline phosphatase) or (B) ß-gal (ß-galactosidase)

- 136 -

Claims

1.   A vaccine comprising a bacterial cell the
virulence of which is attenuated by the constitutive
expression of a gene under the control of a two-component
5   regulatory system.

2.   The vaccine of claim 1, wherein said
constitutive expression is the result of a mutation at a
component of said two-component regulatory system.

3.   The vaccine of claim 1, wherein said
10   bacterial cell comprises a second mutation which
attenuates virulence.

4.   The vaccine of claim 1, wherein said
bacterial cell is a *Salmonella* cell, said two-component
regulatory system is the *phoP* regulatory region, and said
15   gene is a *phoP* regulatory region regulated gene.

5.   The vaccine of claim 4, wherein said
constitutive expression is the result of a mutation.

6.   The vaccine of claim 5, wherein said
mutation is in the *phoP* regulatory region.

20   7.   The vaccine of claim 6, wherein said
mutation is in the *phoP* gene.

8.   The vaccine of claim 6, wherein said
mutation is in the *phoQ* gene.

9.   The vaccine of claim 6, wherein said
25   mutation is a *phoP*^{c} mutation.

- 137 -

10. The vaccine of claim 6, wherein said mutation is a non-revertible mutation.

11. The vaccine of claim 4, wherein said constitutive expression is the result of a change at the promoter of said regulated gene.

12. The vaccine of claim 4, wherein said gene is a *prg* gene.

13. A vaccine comprising a *Salmonella* cell which is attenuated by the decreased expression of a *phoP* regulatory region regulated virulence gene.

14. The vaccine of claim 13, wherein said decrease of expression is the result of a mutation.

15. The vaccine of claim 14, wherein said mutation is in the *prgH* gene.

16. The vaccine of claim 14, wherein said mutation is in the *prgA, prgB, prgC,* or *prgE* genes.

17. The vaccine of claim 4, wherein said gene the is a *pag* gene.

18. The vaccine of claim 17, wherein said *pag* locus is the *pagC* locus.

19. The vaccine of claim 4, further characterized in that said *Salmonella* cell comprises a first mutation which attenuates virulence and a second mutation which attenuates virulence.

- 138 -

20.   The vaccine of claim 19, wherein said first mutation is in a *pho*P regulatory region gene.

21.   The vaccine of claim 20, wherein said first mutation is in the *pho*P gene.

22.   The vaccine of claim 20, wherein said first mutation is in the *pho*Q gene.

23.   The vaccine of claim 20, wherein said first mutation is a *pho*P^c mutation.

24.   The vaccine of claim 19, wherein said first mutation is in a *pho*P regulatory region regulated gene.

25.   The vaccine of claim 19, wherein said second mutation is a mutation in an *aromatic* amino acid synthetic gene.

26.   The vaccine of claim 25, wherein said second mutation is an *aro* mutation.

27.   The vaccine of claim 19, wherein said second mutation is in a *pho*P regulatory region regulated gene.

28.   The vaccine of claim 23, wherein said second mutation is in a *prg* locus.

29.   The vaccine of claim 13, further characterized in that said *Salmonella* cell comprises two mutant genes, a first mutant gene which attenuates virulence and a second mutant gene which attenuates virulence.

- 139 -

30. The vaccine of claim 29, wherein said second gene is in a *prg* locus.

31. The vaccine of claim 30, wherein said gene is *prgH*.

5   32. The vaccine of claim 30, wherein said gene is *prgA*, *prgB*, *prgC*, or *prgE*.

33. The vaccine of claim 27, wherein said second mutation is in a *pag* locus.

34. The vaccine of claim 27, wherein said
10   second mutation is a *pagC* mutation.

35. The vaccine of claim 4, wherein said *Salmonella* is of the species *S. typhi*.

36. The vaccine of claim 4, wherein said *Salmonella* is of the species *S. enteriditis* and of the
15   strain *typhimurium*.

37. The vaccine of claim 4, wherein said *Salmonella* is of the species *S. cholerae-suis*.

38. The vaccine of claim 4, wherein said vaccine is a live vaccine.

20   39. A vaccine comprising a bacterial cell the virulence of which is attenuated by a mutation in a gene under the control of a two-component regulatory system.

40. The vaccine of claim 39, further characterized in that said bacterial cell comprises a
25   virulence attenuating mutation in a second gene.

- 140 -

41. The vaccine of claim 39, wherein said
bacterial cell is *Salmonella* cell and said two-component
regulatory system is the phoP regulatory region.

42. The vaccine of claim 41, wherein said gene
is a *prg* gene.

43. The vaccine of claim 41, wherein said gene
is *prgH*.

44. The vaccine of claim 41, wherein said gene
is *prgA, prgB, prgC,* or *prgE*.

45. The vaccine of claim 41, wherein said gene
is a *pag* gene.

46. The vaccine of claim 45, wherein said gene
is *pagC*.

47. The vaccine of claim 41, wherein said
bacterial cell further comprises a mutation in a second
gene, said mutation attenuating the virulence of said
bacterial cell.

48. The vaccine of claim 47, wherein said
second gene is an *aromatic* amino acid biosynthetic gene.

49. The vaccine of claim 48, wherein said
second gene is an *aro* gene.

50. A vaccine comprising a *Salmonella* cell
comprising a first virulence attenuating mutation in an
aromatic amino acid biosynthetic gene and a second
virulence attenuating mutation in a *phoP* regulatory
region gene.

51. The vaccine of claim 50, wherein said first mutation is at an *aro* gene.

52. The vaccine of claim 51, wherein said second mutation is a phoP⁻ mutation.

5   53. A bacterial cell which constitutively expresses a gene under the control of a two-component regulatory system and which comprises a virulence attenuating mutation which does not result in constitutive expression of a gene under the control of
10   said two-component regulatory system.

54. The bacterial cell of claim 53, further comprising a mutation in a component of said two-component regulatory system.

55. The bacterial cell of claim 53, wherein
15   said cell is a *Salmonella* cell which expresses a *phoP* regulatory region regulated gene constitutively and which comprises a virulence attenuating mutation which does not result in the constitutive expression of a gene under the control of the *phoP* regulatory region.

20   56. The bacterial cell of claim 55, wherein said constitutive expression is caused by a mutation in the *phoP* regulatory region.

57. The bacterial cell of claim 55, wherein said constitutive expression is caused by a mutation in
25   the *phoP* gene.

58. The bacterial cell of claim 55, wherein said constitutive expression is caused by a mutation in the *phoQ* gene.

- 142 -

59.   The bacterial cell of claim 56, wherein said mutation is a *phoP^C* mutation.

60.   The bacterial cell of claim 56, wherein said mutation is a deletion.

5   61.   The bacterial cell of claim 55, further characterized in that said virulence attenuating mutation is a in an *aromatic* amino acid synthetic gene.

62.   The bacterial cell of claim 61, wherein said virulence attenuating mutation is an *aro* mutation.

10   63.   The bacterial cell of claim 55, wherein said virulence attenuating mutation is in a *phoP* regulatory region gene.

64.   The bacterial cell of claim 63, wherein said virulence attenuating mutation is the *phoP* gene.

15   65.   The bacterial cell of claim 63, wherein said virulence attenuating mutation is in the *phoQ* gene.

66.   The bacterial cell of claim 55, wherein said virulence attenuating mutation is in a *prg* locus.

67.   The bacterial cell of claim 66, wherein 20   said virulence attenuating mutation is in the *prgH* gene.

68.   The bacterial cell claim 66, wherein said virulence attenuating mutation is in the *prgA*, *prgB*, *prgC*, or *prgE* gene.

69.   The bacterial cell of claim 55, wherein 25   said virulence attenuating mutation is in a *pag* locus.

- 143 -

70.   The bacterial cell of claim 55, wherein said virulence attenuating mutation is a *pagC* mutation.

71.   The bacterial cell of claim 55, wherein said cell is of the species *S. typhi*.

5   72.   The bacterial cell of claim 55, wherein said cell is of the species *S. enteriditis* and of the strain *typhimurium*.

73.   The bacterial cell of claim 55, wherein said *Salmonella* cell is of the species *S. cholerae-suis*.

10   74.   A bacterial cell comprising a virulence attenuating mutation in a gene regulated by a *phoP* regulatory region.

75.   The bacterial cell of claim 74, wherein said bacterial cell is a *Salmonella* cell and said
15   virulence attenuating mutation is in a *phoP* regulatory region regulated gene.

76.   The bacterial cell of claim 75, wherein said gene is a *prg* gene.

77.   The bacterial cell of claim 76, wherein
20   said gene is the *prgH* gene.

78.   The bacterial cell of claim 76, wherein said gene is the *prgA*, *prgB*, *prgC*, or *prgE* gene.
79.   The bacterial cell of claim 75, wherein said gene is a *pag* gene.

25   80.   The bacterial cell of claim 79, wherein said gene is *pagC*.

- 144 -

81. The bacterial cell of claim 74, further
comprising a second mutation which attenuates virulence
but which does not result in constitutive expression of a
phoP regulatory region regulated gene.

5   82. The bacterial cell of claim 81, wherein
said second mutation is in an *aromatic* amino acid
synthetic gene.

83. The bacterial cell of claim 82, wherein
said second mutation is an *aro* mutation.

10   84. The bacterial cell of claim 81, wherein
said second mutation is in a *phoP* regulatory region gene.

85. The bacterial cell of claim 84, wherein
said second mutation is in the *phoP* locus.

86. The bacterial cell of claim 84, wherein
15   said second mutation is in the *phoQ* locus.

87. The bacterial cell of claim 81, wherein
said second mutation is in a *phoP* regulating region
regulated gene.

88. The bacterial cell of claim 87, wherein
20   said second mutation is in a *pag* locus.

89. The bacterial cell of claim 75, wherein
said cell is of the species *S. typhi*.

90. The bacterial cell of claim 75, wherein
said cell is of the species *S. enteriditis* and of the
25   strain *typhimurium*.

- 145 -

91.   The bacterial cell of claim 75, wherein said cell is of the species *S. cholerae-suis*.

92.   A live *Salmonella* cell in which there is inserted into a virulence gene a gene encoding a heterologous protein, or a regulatory element, of said heterologous protein gene.

93.   The live *Salmonella* cell of claim 92, wherein said virulence gene is in the *phoP* regulatory region.

94.   The live *Salmonella* cell of claim 92, wherein said virulence gene is a phoP regulatory region regulated gene.

95.   The live *Salmonella* cell of claim 94, wherein said virulence gene is a *prg* gene.

96.   The live *Salmonella* cell of claim 95, wherein said virulence gene is the *prgH* gene.

97.   The live *Salmonella* cell of claim 95, wherein said virulence gene is the *prgA*, *prgB*, *prgC*, or *prgE* gene.

98.   The live *Salmonella* cell of claim 94, wherein said virulence gene is a *pag* gene.

99.   The live *Salmonella* cell of claim 98, wherein said *pag* gene is *pagC*.

100.   The live *Salmonella* cell of claim 92, wherein said *Salmonella* cell carries a second mutation that attenuates virulence.

- 146 -

101.   The live *Salmonella* cell of claim 100,
wherein said second mutation is an *aro* mutation.

102.   The live *Salmonella* cell of claim 92,
wherein said DNA encoding a heterologous protein is under
5   the control of an environmentally regulated promoter.

103.   The live *Salmonella* cell of claim 92,
wherein said *Salmonella* cell is of the species *S. typhi.*

104.   The live *Salmonella* cell of claim 92,
further comprising a DNA sequence encoding T7 polymerase
10   under the control of an environmentally regulated
promoter and a T7 transcriptionally sensitive promoter,
said T7 transcriptionally sensitive promoter controlling
the expression of said heterologous antigen.

105.   A vector capable of integrating into the
15   chromosome of *Salmonella* comprising
a first DNA sequence encoding a heterologous
protein,
a second DNA sequence encoding a marker, and
a third DNA sequence encoding a *phoP* regulatory
20   region regulated gene product necessary for virulence,
said third DNA sequence being mutationally inactivated.

106.   The vector of claim 105, wherein said
phoP regulatory region regulated gene is a *prg* locus.

107.   The vector of claim 106, wherein said
25   gene is *prgH.*

108.   The vector of claim 106, wherin said gene
is *prgA, prgB, prgC,* or *prgE.*

- 147 -

109. The vector of claim 105, wherein said *phoP* regulatory region regulated gene is a *pag* locus.

110. The vector of claim 109, wherein said *pag* locus is *pagC*.

5   111. The vector of claim 105, wherein said first DNA sequence is disposed on said vector so as to mutationally inactivate said third DNA sequence.

112. The vector of claim 105, wherein said vector cannot replicate in a wild-type *Salmonella* strain.

10   113. The vector of claim 105, wherein said first DNA sequence encoding a heterologous protein is under the control of an environmentally regulated promoter.

114. The vector of claim 105, further
15 comprising a DNA sequence encoding T7 polymerase under the control of an environmentally regulated promoter and a T7 transcriptionally sensitive promoter, said T7 transcriptionally sensitive promoter controlling the expression of said first DNA sequence encoding a
20 heterologous protein.

115. A method of vaccinating an animal against a disease caused by a bacterium comprising administering the vaccine of claim 1.

116. The method of claim 115, wherein said
25 bacterium is *Salmonella* and said vaccine is the vaccine of claim 4.

- 148 -

117. A method of vaccinating an animal against a disease caused by a bacterium comprising administering the vaccine of claim 39.

118. The method of claim 115, wherein said bacterium is *Salmonella* and said vaccine is the vaccine of claim 41.

119. A method of vaccinating an animal against a disease caused by *Salmonella* comprising administering the vaccine of claim 50.

120. A vector comprising DNA which encodes the *pagC* gene product.

121. A cell comprising the vector of claim 120.

122. A method of producing the *pagC* gene product comprising culturing the cell of claim 121 and purifying the *pagC* gene product from said cell or culture medium.

123. A purified preparation of the *pagC* gene product.

124. A method of detecting the presence of *Salmonella* in a sample comprising contacting said sample with *pagC* encoding DNA and detecting the hybridization of said *pagC* encoding DNA to nucleic acid in said sample.

125. A vector comprising DNA which encodes the *prgH* gene product.

- 149 -

126. A cell comprising the vector of claim 125.

127. A method of producing the *prgH* gene product comprising culturing the cell of claim 126 and
5   purifying the *prgH* gene product from said cell or culture medium.

128. A purified preparation of the *prgH* gene product.

129. A method of detecting the presence of
10 *Salmonella* in a sample comprising contacting said sample with *prgH* encoding DNA and detecting the hybridization of said *prgH* encoding DNA to nucleic acid in said sample.

130. A method of attenuating the virulence of a bacterium, said bacterium comprising a two-component
15 regulatory system, comprising causing a gene under the control of said two-component system to be expressed constitutively.

131. The method of claim 124, wherein said bacterium is *Salmonella* and said two-component system is
20 the *phoP* regulatory region.

132. A bacterial cell the virulence of which is attenuated by a first mutation in a PhoP regulon and a second mutation in an aromatic amino acid synthetic gene.

133. The bacterial cell of claim 132, wherein
25 said bacterial cell is a *Salmonella* cell.

- 150 -

134. The *Salmonella* cell of claim 133, wherein said *Salmonella* cell is a *Salmonella typhimurium* cell.

135. The *Salmonella* cell of claim 133, wherein said *Salmonella* cell is a *Salmonella enteriditis*.

5   136. The *Salmonella* cell of claim 135, wherein said *Salmonella* cell is a *Salmonella pylorum* cell.

137. The *Salmonella* cell of claim 135, wherein said *Salmonella* cell is a *Salmonella paratyphi A* cell.

138. The *Salmonella* cell of claim 135, wherein
10   said *Salmonella* cell is a *Salmonella paratyphi B* cell.

139. The *Salmonella* cell of claim 133, wherein said *Salmonella* cell is a *Salmonella cholerasuis* cell.

140. The *Salmonella* cell of claim 133, wherein said *Salmonella* cell is a *Salmonella typhi* cell.

15   141. The bacterial cell of claim 133, wherein said first mutation comprises a non-revertable null mutation in the PhoP/PhoQ locus.

142. The bacterial cell of claim 141, wherein said mutation comprises a deletion of at least 100
20   nucleotides.

143. The bacterial cell of claim 142, wherein said mutation comprises a deletion of at least 500 nucleotides.

- 151 -

144. The bacterial cell of claim 143, wherein said mutation comprises a deletion of at least 750 nucleotides.

145. The bacterial cell of claim 144, wherein said mutation comprises a deletion of nucleotides 376 to 1322 of said *PhoP/PhoQ* locus.

146. The bacterial cell of claim 141, wherein said second mutation comprises a non-revertable null mutation in an AroA locus.

147. The bacterial cell of claim 141, wherein said second mutation comprises a non-revertable null mutation in an AroC/AroD locus.

148. The bacterial cell of claim 146, further comprising a mutation in a non-aromatic amino acid synthetic gene, wherein said mutation renders said cell auxotrophic for said non-aromatic amino acid.

149. The bacterial cell of claim 148, wherein said amino acid is histidine.

150. The bacterial cell of claim 149, wherein said *S. typhi* has the genotype AroA⁻, His⁻, PhoP/PhoQ⁻.

151. The bacterial cell of claim 150, wherein said *S. typhi* is TyH445.

152. The bacterial cell of claim 134, wherein wherein said first mutation comprises a non-revertable null mutation in the PhoP/PhoQ locus.

- 152 -

153. The bacterial cell of claim 152, wherein said mutation comprises a deletion of nucleotides 376-1322 of said *PhoP/PhoQ* locus.

154. The bacterial cell of claim 152, wherein said second mutation comprises a non-revertible null mutation in an AroA locus.

155. The bacterial cell of claim 154, further comprising a mutation in a non-aromatic amino acid synthetic gene, wherein said mutation renders said cell auxotrophic for said non-aromatic amino acid.

156. A vaccine comprising the bacterial cell of claim 132.

157. A substantially pure DNA comprising a sequence encoding *pagD*.

158. The DNA of claim 157, wherein said sequence comprises nucleotides 91 to 354 of SEQ ID NO:5.

159. The DNA of claim 158, further comprising nucleotides 4 to 814 of SEQ ID NO:15.

160. A substantially pure DNA comprising nucleotides 4 to 814 of SEQ ID NO:15.

161. The DNA of claim 160, wherein said DNA sequence comprises nucleotides 562 to 814 of SEQ ID NO:15.

162. The DNA of claim 160, wherein said DNA sequence comprises nucleotides 4 to 776 of SEQ ID NO:15.

- 153 -

163. The DNA of claim 158 and degenerate variants thereof, wherein said sequence encodes a product which comprises essentially the amino acid sequence given in SEQ ID NO:6.

5   164. A substantially pure DNA comprising a sequence encoding envE.

165. The DNA of claim 164, wherein said sequence comprises nucleotides 1114 to 1650 of SEQ ID NO:5.

10   166. The DNA of claim 165 and degenerate variants thereof, wherein said sequence encodes a product which comprises essentially the amino acid sequence given in SEQ ID NO:7.

167. A substantially pure DNA comprising a
15 sequence encoding msgA.

168. The DNA of claim 167, wherein said sequence comprises nucleotides 1825 to 2064 of SEQ ID NO:5.

169. The DNA of claim 168, further comprising
20 nucleotides 1510 to 1824 of SEQ ID NO:5.

170. A substantially pure DNA comprising nucleotides 1510 to 1760 of SEQ ID NO:5.

171. The DNA of claim 168 and degenerate variants thereof, wherein said sequence encodes a product
25 which comprises essentially the amino acid sequence given in SEQ ID NO:8.

- 154 -

172.   A substantially pure DNA comprising a
sequence encoding *envF*.

173.   The DNA of claim 172, wherein said
sequence comprises nucleotides 2554 to 3294 of SEQ ID
5   NO:5.

174.   The DNA of claim 173, further comprising
nucleotides 2304 to 2553 of SEQ ID NO:5.

175.   The DNA of claim 173 and degenerate
variants thereof, wherein said sequence encodes a product
10   which comprises essentially the amino acid sequence given
in SEQ ID NO:9.

176.   A substantially pure DNA comprising the
sequence given in SEQ ID NO:5 or a fragment thereof.

177.   A substantially pure DNA comprising the
15   sequence given in SEQ ID NO:10 or a fragment thereof.

178.   A substantially pure DNA comprising a
sequence encoding *prgH*.

179.   The DNA of claim 178, wherein said
sequence comprises nucleotides 688 to 1866 of SEQ ID
20   NO:10.

180.   The DNA of claim 179, further comprising
nucleotides 1 to 689 of SEQ ID NO:10.

181.   The DNA of claim 179 and degenerate
variants thereof, wherein said sequence encodes a product
25   which comprises essentially the amino acid sequence given
in SEQ ID NO:11.

- 155 -

182. A substantially pure DNA comprising a sequence encoding *prgI*.

183. The DNA of claim 182, wherein said sequence comprises nucleotides 1891 to 2133 of SEQ ID NO:10.

184. The DNA of claim 183, further comprising nucleotides 1 to 689 of SEQ ID NO:10.

185. The DNA of claim 183 and degenerate variants thereof, wherein said sequence encodes a product which comprises essentially the amino acid sequence given in SEQ ID NO:12.

186. A substantially pure DNA comprising a sequence encoding *prgJ*.

187. The DNA of claim 186, wherein said sequence comprises nucleotides 2152 to 2457 of SEQ ID NO:10.

188. The DNA of claim 187, further comprising nucleotides 1 to 689 of SEQ ID NO:10.

189. The DNA of claim 187 and degenerate variants thereof, wherein said sequence encodes a product which comprises essentially the amino acid sequence given in SEQ ID NO:13.

190. A substantially pure DNA comprising a sequence encoding *prgK*.

- 156 -

191.   The DNA of claim 190, wherein said
sequence comprises nucleotides 2456 to 3212 of SEQ ID
NO:10.

192.   The DNA of claim 191, further comprising
nucleotides 1 to 689 of SEQ ID NO:10.

193.   The DNA of claim 191 and degenerate
variants thereof, wherein said sequence encodes a product
which comprises essentially the amino acid sequence given
in SEQ ID NO:14.

194.   A bacterial cell, the virulence of which
is attenuated by a mutation in one or more genes selected
from the group consisting of *pagD*, *pagE*, *pagF*, *pagG*,
pagH, *pagI*, *pagJ*, *pagK*, *pagL*, *pagM*, *pagN*, *pagP*, *envE*, and
envF.

195.   A bacterial cell, the virulence of which
is attenuated by a mutation in one or more genes selected
from the group consisting of *pagC*, *pagD*, *pagJ*, *pagK*, *pagM*,
and *msgA*.

196.   A bacterial cell, the virulence of which
is attenuated by a mutation in one or more genes selected
from the group consisting of *prgH*, *prgI*, *prgJ*, and *prgK*.
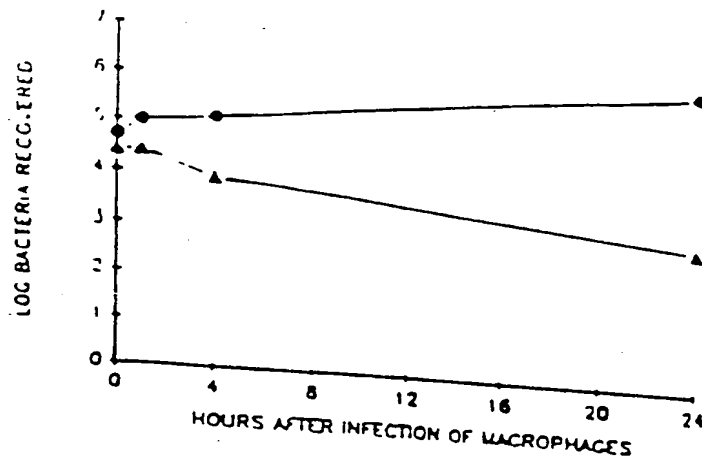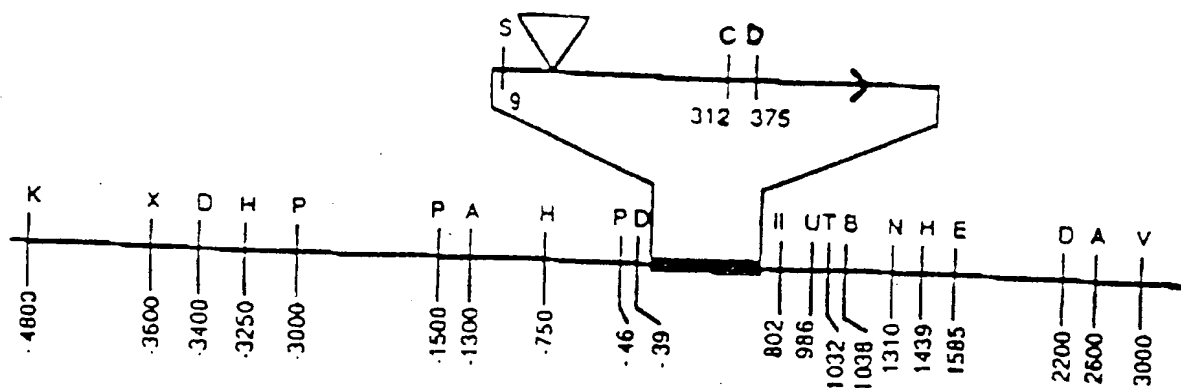
FIGURE 1

FIGURE 2

```
        10         20         30         40         50         60         70
GTTAACCACT CTTAATAATA ATGGGTTTTA TAGCGAAATA TACTTTTTTA TCGCGTGTTC AATATTTCCG

        80         90        100        110        120        130      --140
TTAGTTATTA TTTTTTTGGA ATCTAAATTC TCTCTAAACA CAGGTCATAT TTATGTTGGA ATTGTGGTGT

        150        160        170        180        190        200        210
TGATTCTATT CTTATAATAT AACAACAAAT GTTGTAACTG ATAGATATAT TAAAAGATTA AATCCGAGCG
                             * ────────►

        220        230        240        250        260        270        280
GGAATAAAGC CTGCTAAGCA TCATCCTGAA TATGATTACA GCGCCTGCGA TGGCATATAA CCGTATTGCG

        290        300        310        320        330        340        350
GATCGAGCGT CACGTGAGCA CTGTGAAGCA CAATGCGATA TGTTCTCATT ATATGGCGAG TTTGCTTAAT

        360        370        380        390        400        410        420
CACATGTTTT TAGCCGAACG GTGTCAAGTT TCTTAATGTG GTTGTGAGAT TTTCTCTTTA AATATCAAAA

        430        440        450        460        470        480        490
TGTTCCATGG GTGATTTGTT GTTCTATAGT CGCTAAACAC TTTATGGTTT CTGTTAAATA TATATGCGTC

        500        510        520        530        540        550        560
AGAAAAATTA GCATTCAAAT CTATAAAAGT TAGATGACAT TGTAGAACCC GTTACCTAAA TGAGCCATAC

        570        580        590        600        610        620        630
AGTCCTTCGG TAGTAAAAAT ATCTTTCAGG AAGTAAACAC ATCAGCAGCG ATAGCGGTGA ATTATTCGTC

        640        650        660        670        680        690        700
GTTTTGTCGA TTCCCCATAG TCGCGATAAC TGAATGCCCG ATCGGTACTG CAGGTGTTTA AACACACCGT

        710        720        728
AAATAATAAG TAGTATTAAG GAGTTGTT
```

```
ATG AAA AAT ATT ATT TTA TCC ACT TTA GTT ATT ACT ACA AGC GTT TTG GTT GTA   782
MET LYS ASN ILE ILE LEU SER THR LEU VAL ILE THR THR SER VAL LEU VAL VAL    18

AAT GTT GCA CAG GCC GAT ACT AAC GCC TTT TCC GTG GGG TAT GCA CGG TAT GCA   836
ASN VAL ALA GLN ALA ASP THR ASN ALA PHE SER VAL GLY TYR ALA ARG TYR ALA    36

CAA AGT AAA GTT CAG GAT TTC AAA AAT ATC CGA GGG GTA AAT GTG AAA TAC CGT   890
GLN SER LYS VAL GLN ASP PHE LYS ASN ILE ARG GLY VAL ASN VAL LYS TYR ARG    54

TAT GAG GAT GAC TCT CCG GTA AGT TTT ATT TCC TCC CTA AGT TAC TTA TAT GGA   944
TYR GLU ASP ASP SER PRO VAL SER PHE ILE SER SER LEU SER TYR LEU TYR GLY    72

GAC AGA CAG GCT TCC GGG TCT GTT GAG CCT GAA GGT ATT CAT TAC CAT GAC AAG   998
ASP ARG GLN ALA SER GLY SER VAL GLU PRO GLU GLY ILE HIS TYR HIS ASP LYS    90

TTT GAG GTG AAC TAC GGT TCT TTA ATG GTT GGG CCA GCC TAT CGA TTG TCT GAC  1052
PHE GLU VAL LYS TYR GLY SER LEU MET VAL GLY PRO ALA TYR ARG LEU SER ASP   108

AAT TTT TCG TTA TAC GCG CTC GCG GGT GTC GGC ACG GTA AAG GCG ACA TTT AAA  1106
ASN PHE SER LEU TYR ALA LEU ALA GLY VAL GLY THR VAL LYS ALA THR PHE LYS   126

GAA CAT TCC ACT CAG GAT GGC GAT TCT TTT TCT AAC AAA ATT TCC TCA AGG AAA  1160
GLU HIS SER THR GLN ASP GLY ASP SER PHE SER ASN LYS ILE SER SER ARG LYS   144

ACG GGA TTT GCC TGG GGC GCG GGT GTA CAG ATC AAT CCC CTG GAG AAT ATC GTC  1214
THR GLY PHE ALA TRP GLY ALA GLY VAL GLN MET ASN PRO LEU GLU ASN ILE VAL   162
```

FIGURE 3 page 1 of 2

```
GTC GAT GTT GGG ⁻ ⁻ GAA GGA AGC AAC ATC TCC TCT A( \AA ATA AAC GGC TTC 1268
VAL ASP VAL GLY ⁚ ⹁ GLU GLY SER ASN ILE SER SER TH. LYS ILE ASN GLY PHE 180

AAC GTC GGG GTT GGA TAC CGT TTC TGA AAAGC                           ⸺ 1300
ASN VAL GLY VAL GLY TYR ARG PHE                                    ⸺ 188


       1310      1320      1330      1340      1350      1360      1370
ATAAGCTATG CGGAAGGTTC GCCTTCCGCA CCGCCAGTCA ATAAAACAGG GCTTCTTTAC CAGTGACACG


       1380      1390      1400      1410      1420      1430      1440
TACCTGCCTG TCTTTTCTCT CTTCGTCATA CTCTCTTCGT CATAGTGACG CTGTACATAA CATCTCACTA


       1450      1460      1470      1480      1490      1500      1510
GGATAAGCAC AGATAAAGGA TTGTGGTAAG CAATCAAGGT TGCTCAGGTA GGTGATAAGC AGGAAGGAAA


       1520      1530      1540      1550      1560      1570      1580
ATCTGGTGTA AATAACGCCA GATCTCACAA GATTCACTCT GAAAAATTTT CCTGGAATTA ATCACAATGT


       1590      1600      1610      1620      1630      1640      1650
CATCAAGATT TTGTCACCGC CTTCGCATAT TGTACCTGCCG CTGAACGAC TACTGAAAAG TAGGCAAGGTA


       1660      1670      1680      1690      1700      1710      1720
TGTATTTTAT CCAGGAGAGC ACCTTTTTTG CGCCTGGCAG AAGTCCCCAG CCGCCACTAG CTCAGGTCGA


       1730      1740      1750      1760      1770      1780      1790
TAGAGCATCA ACCTCCTAA GTTGATGGTGC GAGGTTCGAG GCCTCGGTGG CGGTCCAATG TGGTTATCGT


       1800      1810      1820      1830      1840      1850      1860
ATAATGTTAT TACCTCAGT GTCAGGCTGAT GATGTGGGTT CGACTCCCAC TGACCACTTC AGTTTTGAAT


       1870      1880      1890      1900      1910      1920      1930
AAGTATTCTC TCCCAACCC TGTTACAGAAT AATTTCATTT ATTACGTGAC AAGATAGTCA TTTATAAAAA


       1940      1950      1960      1970      1980      1990      2000
ATGCACAAAA ATGTTATTG TCTTTTATTAC TTGTGAGTTG TAGATTTTTC TTATGCGGTG AATCCCCCTT


       2010      2020      2030      2040      .2050     2060      2070
TGCGGCGGGC CGTCCACTC AAATAGTTAAT GTTCCTCGCG AACCATATTG ACTGTGGTAT GGTTCACCGG


       2080      2090      2100      2110      2120      2130      2140
GAGGCACCCG GCACCGCAA TTTTTTATAAA ATGAAATTCA CACCCTATGG TTCAGAGCGG-TCTCTTTTTA


       2150      2160      2170      2180      2190      2200      2210
CATCAGGTGG GCAAGCATA ATGCAGGTTAA CTTGAAAGAT ACGATCAATA GCAGAAACCA GTGATTTCGT


       2220      2230      2240      2250      2260      2270      2280
TTATGGCCTG GGGATTTAA CCGCGCCAGAG CGTATGCAAG ACCCTGCCGC GGTTGGCCGG TGATCGTTCA


       2290      2300      2310
ATAGTGCGAA TATGAATGG TTACCAGCCGC TGCGAATTC (Sequence I.D. No.1)
```

FIGURE 3 page 2 of 2

FIGURE 4

```
5'        GAG  CGC  ATT  ATC  AGA  TAA  ATT  GAT  TTA  TTTCTCACT

TTC       ATT  CTA  TTT  TCA      TCA

GGA       ATC  CCT  GTG  TCC  TGT  GCG  GTA  ATC  TGC  TGCTATCGA

GAA       CGA  CAG  ACA  TCG

CTA       ACA  GTA  TAT  ATG  GAA  ACA  TCA  AAA  GAG  AAGACGATA

ACA       AGC  CCA  GGG  CCA  TAC

ATA       GTT  CGA  TTA  CTT  AAC  AGC  TCA  CTG  AAC  GGCTGTGAG

TTT       CCA  TTG  CTG  ACA  GGC

CGA       ACA  CTC  TTT  GTG  GTA  GGT  CAG  AGT  GAT  GCGCTCACT

GCT       TCA  GGT  CAA  CTC  CCT

GAT       ATA  CCT  GCC  GAT  AGC  TTT  TTT  ATC  CCG  CTGGACCAT

GGC       GGA  GTA  AAT  TTT  GAA

ATC       CAG  GTG  GAT  ACG  GAT  GCG  ACC  GAA  ATT  ATACTCCAT

GAG       CTG  AAA  GAA  GGA  AAT

TCT       GAA  TCT  CGT  TCG  GTG  CAA  TTA  AAT  ACG  CCAATACAG

GTC       GGT  GAA  TTG  CTT  ATC

CTG       ATT  CGC  CCG  GAA  AGC  GAG  CCG  TGG  GTG  CCCGAGCAG

CCT       GAG  AAG  TTA  GAA  ACG

TCT       GCA  AAA  AAG  AAC  GAG  CCG  CGT  TTT  AAA  AACGGAATT

GTA       GCA  GCA  CTG  GCC
```

FIGURE 5 page 1 of 2

```
GGG     TTT TTT ATA TTG GGA ATT GGG ACT GTG GGGACGTTA

TGG     ATA CTT AAC TCG CCG

CAG     CGG CAG GCC CGA GAG CTC GAT TCG TTA TTGGGGCAG

GAG     AAG GAG CGT TTT CAG GTG     TTG CCA GGCC 3'
```
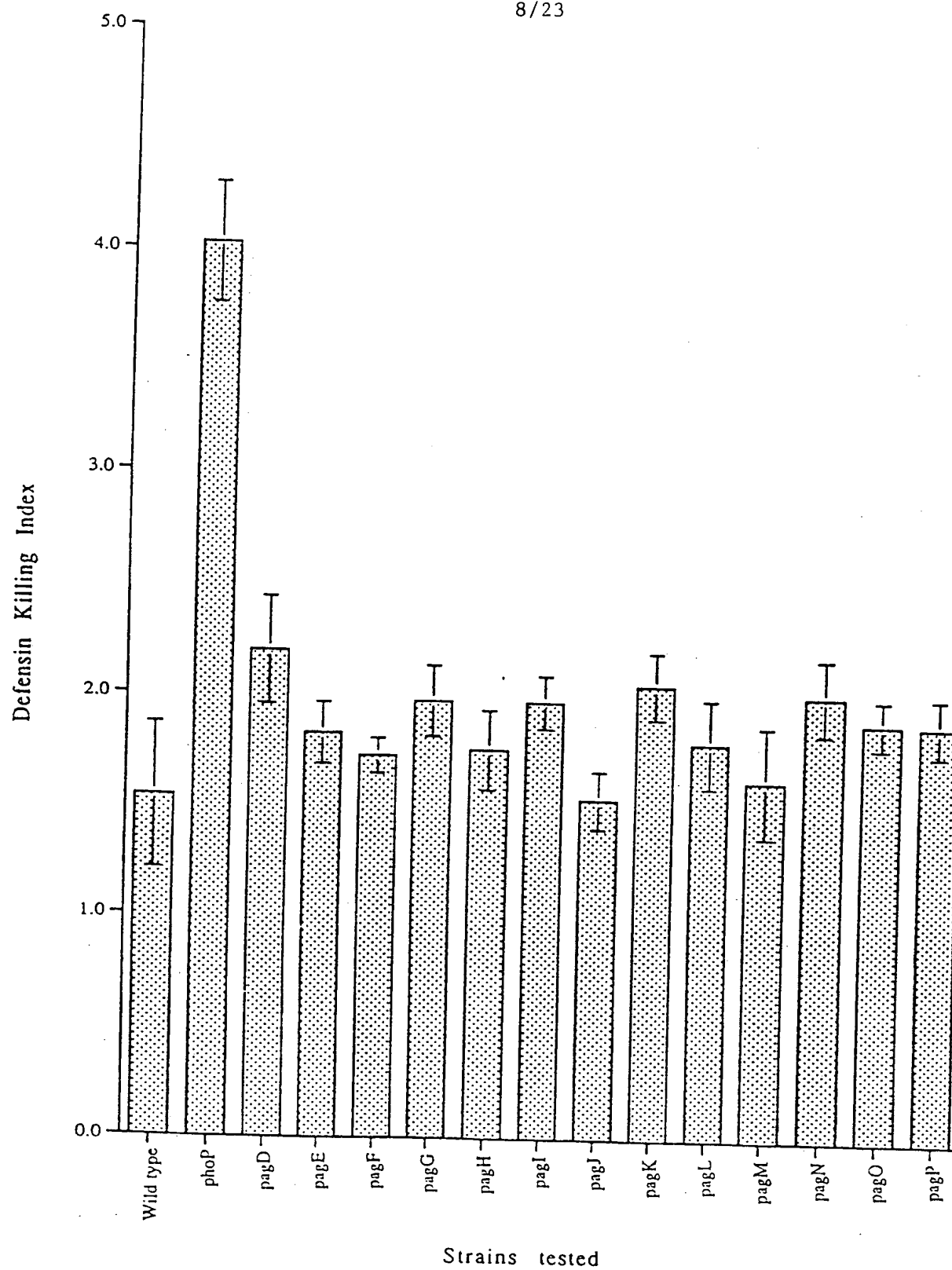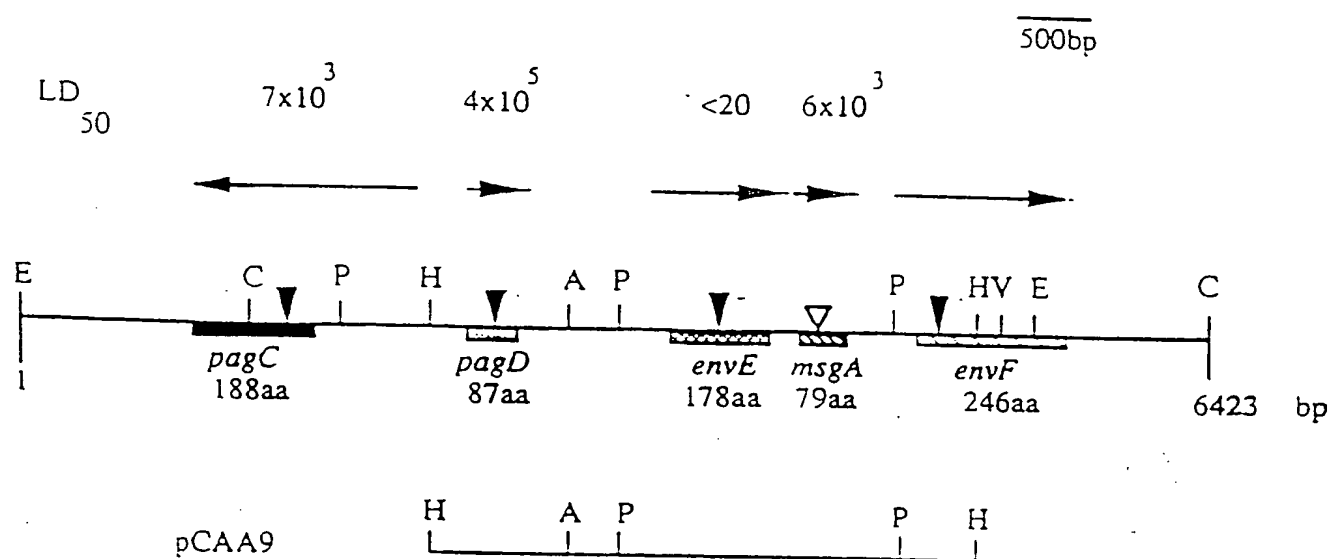
FIGURE 5 page 2 of 2

FIGURE 6

FIGURE 7

```
    _HpaI_      ······                        ······         —
  1 GGTTAACTCTTCGTTGAATAAAAAATGTCAATGACGTTCCATAATTCAGGACATGAACTT   60
 61 CACAAGTCATTATATATAACAGGAGGTGCTATGAAACATCATGCTTTTATGCTTTGGTCA  120
                              M  K  H  H  A  F  M  L  W  S
                          pagD        ▼
121 TTACTTATTTTTTCATTCCATGTTTTGGCCAGTTCAGGCCATTGTTCTGGTTTACAACAG  180
     L  L  I  F  S  F  H  V  L  A  S  S  G  H  C  S  G  L  Q  Q
181 GCATCATGGATATTTTTTATCTACGATTTTGGTAGTAAAACCCCGCAACCACCTACAAAT  240
     A  S  W  D  I  F  I  Y  D  F  G  S  K  T  P  Q  P  P  T  N
241 ACTGATAAAAAGCAAGCCAGGCAGATTAGTTCACCGTCCTGCCCGACGACAAAACCCATG  300
     T  D  K  K  Q  A  R  Q  I  S  S  P  S  C  P  T  T  K  P  M
301 ATGTCCCCACCAGTCAATGACGCCAGGAAAGGGAATACTTTCTCCAGAACATAATGTTAT  360
     M  S  A  P  V  N  D  A  R  K  G  N  T  F  S  R  T  *    (SEQ  ID  NO:6)
361 TTATCTACAATGGTGCCGACGACTACTTTTAGCCACCCGGAAATCTTGATTGCCATCAAA  420
421 TATAGCTGGCATTATTTTTCCTGACGTGTATAGTGCCGCCTCGTTATCCCCATTAAGGAAT  480
481 TTGTTTCTCTCGTAAAATGACAGGAATTGTCAAAACCTTTGATTGTAAGAGCGGTAAAGG  540
541 TCTCATCACCCCCTCCGATGACGCAAAGATGTTCAGGTCCACATTTCAGCATGTCGCCAA  600
601 CACGAAACAGAAGCGCTTATCCCCCGGTATACGCGTTGAGTTTTATCGTATTAATGGCCTC  660
661 CGCCGGACCTACCGCCGCCAACGTTTATCTTTCATAATTCGTCACCCGGCATTTTTCAGAA  720
721 AAATTTAGCGAGTACGTCTACCTCCGCAGCCTCCTATGAGGCTTTGCCTGAAAGGCTGCA  780
781 GAATGTTTTCAGTGGCGAAAATCTAAAAGATTTATTTTCCTAATCAGTCCTGTGACCTCT  840
841 TTTATCATATATCGGGTGCCCCCCCCTTCTCACTTTGTTTAACGTGAAGAAATGTACAGCC  900
901 CTTTTTCACTGTGATAGCATCTAATATTGCAAAAGTATTTAACGCTATATACCCATTGTC  960
961 ACAGGAGTGGCTGCGTGCGAGCTGAGCTATTTAACCGAAGTATTTATGTGATCATTGGAA 1020
1021 TTATCTCTATTGCCGCTCAATGCTACGTCATATTCAGTGGGTATAAATCGCCAATATAGT 1080
1081 TGTAACGCTATTTATTTTTAGGGTAATAATTGAATGACTTTGCTTTCACGAAAAACCACA 1140
                         M  T  L  L  S  G  K  T  T
                         envE
1141 CTCGTTCTCTGCCTCTCGTCTATTTTATGTCGATGTACGACGAACGGCTTACCCACACCT 1200
     L  V  L  C  L  S  S  I  L  C  G  C  T  T  N  G  L  P  T  P
1201 TATAGTATTAATTTCTCGTTCCCCGGTCATTACACAAAACCAGATTAATTCCGGTCGTTAT 1260
     Y  S  I  N  L  S  F  P  V  I  T  Q  N  Q  I  N  S  G  C  Y
1261 TACATAAATGACGCGGAACAAATTCGGACAACTGATGGTCTCTGCCTTCATGCACGCCCA 1320
     Y  I  N  D  A  E  Q  I  R  T  T  D  G  L  C  L  D  A  G  P
1321 GATCAACAGAATCGTTTCACGCTCCGGGAGTGTAAGCATGTGCAATCTCAGCTTTTCTCA 1380
     D  Q  Q  N  R  L  T  L  R  E  C  K  H  V  Q  S  Q  L  F  S
```

FIGURE 8 page 1 of 3

11/23

```
1381  TTTCACCGAGACAGAATCACGCACGGTGAGAAATGTCTGGATGCCCCAGACAAGGTACAA  1440
      F  H  R  D  R  I  T  Q  G  E  K  C  L  D  A  A  D  K  V  Q

1441  AAGAAGGCACACCAATCATTCTTTATTCATGCACGGGTAATGATAACCAGCGCTGGCTCA  1500
      K  K  A  H  Q  S  F  F  I  H  A  R  V  M  I  T  S  A  G  S

1501  CTGATCATAACAAAATTAAGGGGAAACAGAGCCCGAAAATGCCTGGGCACAAATAGCATTA  1560
      L  I  I  T  K  L  R  G  N  R  A  E  N  A  W  A  Q  I  A  L

1561  TTGTCAGAAAAGGCGACCCTGTTTGTGTTGGCCCGATTGCGATTTTTAGTCGCGCCCTGGAAT  1620
      L  S  E  K  A  T  L  L  C  W  P  I  A  I  L  V  A  P  W  N

1521  TTACCATCAGGTAGCAGGACACCCCTGTGAAGAGAGTGCCGGCTAACTTTTCAACACGACA  1580
      L  P  S  G  S  R  T  P  L  ·  (SEQ ID NO:7) ·

1681  ACAGGTTAGCGACCTTTACTTCCACGTGCGATCAATTTACTTTACGTCCGCAACGTCACG  1740

1741  ATGACAAAACGGCGGCTAAACCTTGACACCAGTTATATACCCAGCTTAAATACTGGTCAT  1800

1801  CCAACCAGTAAAAAGGAAATGGCGATGTTCGTCGAACTCGTTTATGACAAGCGAAATGTT  1860
                     M  F  V  E  L  V  Y  D  K  R  N  V
                     msgA

1861  GAAGGTTTGCCAGGCGCACGCGAAATCATCCTCAATGAACTCACAAAACGCGTACATCAA  1920
      E  G  L  P  G  A  R  E  I  I  L  N  E  L  T  K  R  V  H  Q

1921  CTTTTTCCCGATGCGCAAGTGAAAGTTAAGCCAATGCAGGCGAACGCATTAAACAGTGAC  1980
      L  F  P  D  A  Q  V  K  V  K  P  M  Q  A  N  A  L  N  S  D

1981  TGTACAAAAACCGAGAAAGAACGGCTCCACCGTATGCTGGAAGAGATGTTTGAAGAGGCT  2040
      C  T  K  T  E  K  E  R  L  H  R  M  L  E  E  M  F  E  E  A

2041  GATATGTGGCTCGTCGCCCGAATAACGTCCCCTCCTGCGAAAGCCAACATGTCCGATCGAA  2100
      D  M  W  L  V  A  E  ·  (SEQ ID NO:8) ·

2101  AACACCGGGCCTGAGGCGGCTGTCTGTCACGATATAACGCAAACGCTACCACTCAGAACATG  2160

2161  TTGTTGTTGATACCTCAGACCGGTATCTGGAACCGACATTCATCGCTTCACTGGCCTGTC  2220

2221  CGTATGAGTAGCCCTTATCAACAATCAGCTCTGCGCATTCCAGCCTGAAATCTGAAGTA  2280

2281  CGTTTGGTTTTGTTGTTTATTAAGAGCCTATCCCATTAGACTCTTTTATTCGCCAAACTG  2340

2341  GCTTTAACGATTACGCCTACTCGGATAGGTTCTAAACTTATCATCAATACGTAAAATACC  2400

2401  TATTTACGAACAAAAAGTAACAGGTAAAAATCCGAAATAAAACCAGCATAACTAAAACTT  2460

2461  ACTGCAGATATGCACACGCATTATTACTATGTTTCCAGGATAGTCTGGACCAGTCAAGAC  2520

2521  TATCTATTTTATATAAAAAGGGAAATACTTCACATGAATAAAATACATGTTACATATAAA  2580
                                          M  N  K  I  H  V  T  Y  K
                                          envF

2581  AATCTCTTACTTCCGATTACCTTCATCGCCGCCAACTCTAATTAGCGCCTGTGATAACGAT  2640
      N  L  L  L  P  I  T  F  I  A  A  T  L  [I  S  A  C]  D  N  D

2641  AAAGATGCCATGGCGGAAGCTGAAAAAAATCAAGAGAAATACATGCAAAAAATCCAGCAA  2700
      K  D  A  M  A  E  A  E  K  N  Q  E  K  Y  M  Q  K  I  Q  Q
```

FIGURE 8 page 2 of 3

```
2701  AAAGAGCACCACCAATCAATCTTCTTTTACGACAAAGCCCAAATCCAAAAAGCTATTCCC   2760
      K  E  H  Q  Q  S  M  F  F  Y  D  K  A  E  M  Q  K  A  I  A

2761  AATATCAACGCAAAAGGTGGAGCCAATCTTCCGATTATTGAAGTCCGTTTCTTCAAGGGC   2820
      N  I  N  A  K  G  G  A  N  L  A  I  I  E  V  R  F  F  K  G

2821  CGGTATTCATTCATTCGACAAAGCGTTAACACCCCTGCTAAAGTAGACGGTGTTTAAATTT   2880
      G  Y  S  F  I  R  Q  S  V  N  T  P  A  K  V  E  V  F  K  F

2881  AACAACGGCTACTGGGGGGGACCTTCGCCTGTCAATTTAACCATCTTTGGCACTATAACA   2940
      N  N  G  Y  W  G  G  P  S  P  V  N  L  T  I  F  G  T  I  T

2941  GAGGAGCAAAAACAAGAAGCACTAAAAGAGGCTTTATTCAAATTCGACTCGATCAATTTC   3000
      E  E  Q  K  Q  E  A  L  K  E  A  L  F  K  F  D  S  I  N  F

3001  AGCATTATACCAGAGCGTATTCAGGAAACAATTAAACGCGCTAACGCCAGTGGCATCATT   3060
      S  I  I  P  E  R  I  Q  E  T  I  K  R  A  N  A  S  G  I  I

3061  TCCGTTACGGAAGATAGCGATATCGTTGTACGAGCAGAGATAGCTCATAATGGCGAATTC   3120
      S  V  T  E  D  S  D  I  V  V  R  A  E  I  A  H  N  G  E  F

3121  GTCTATGACATTACCATCACTGCTAAAAATACAGCACGTGCGGTAATGACCTTAAATAAG   3180
      V  Y  D  I  T  I  T  A  K  N  T  A  R  A  V  M  T  L  N  K

3181  GATGGTTCTATTGCCGGATATGAGATCAAAGAACCTTTCGCCCCAAAAAAAGAAGCCGAA   3240
      D  G  S  I  A  G  Y  E  I  K  E  P  F  A  P  K  K  E  A  E

3241  AAAGCACAGCAACTTCTTGAACAATCGAGAAAAGACATTGAAAGTCCAGCGTAAAAAAGC   3300
      K  A  Q  Q  L  V  E  Q  S  R  K  D  I  E  S  P  A  · (SEQ ID NO:9)

3301  AGCTCGAAAGATCAACGAAATACAGCAGACATTTAAAAATAGCAGGCGATACAAACATTC   3360

3361  ATAAAAATTATAGCCGCGAAAGAGCGCGTGCCACGTACTAAGGCACTGCTTGAAGACAGCG   3420

3421  AATCGCTATTTCATTCTCTGACACTGTAATTTTTCGTACTCAAGATGTTTATTTATTGAG   3480

3481  TCTTTTCTGGATAACCAGGTGAAGTTATGTCACGCCAGGAATCTATTCCAGCGGGCGTAC   3540

3541  TTGTTGGAGCCAGTGTGAAGCCCGGCAGCGCGCAGAAACCGGAGCGTATACGTTGTACGT   3600

3601  AAGAATTTCGAGCCACTGCCCCACCTAAAAATCATGAATAAAATAGATATTTTAAAGAGGT   3660

3661  AATATCAAGAATTTTTTTCAAAATAATTACTGATTTCATCGCCGATATTTCCCTTCATCTA   3720

3721  TTTCCTATATTTTTATGCATGTTATTCGTATACAAAACAGGACCATCAATTGGTGTCATA   3780

3781  TCATTTTTTTATTCCATTAATTATTTATATCATTCTTCATTTTTTTTTTACTCATTTCTTCA   3840

3841  AAAAATCATAAAAAAAAATATTCAAATAAGTATTTAAAATTATTGTTTTCTGGTACAAATT   3900

3901  CAGCGCAATAAAACAGAGCAACTAAAAAAAATTAGGCGTAGCGAAGTGGAAAAGGACTGT   3960

3961  CATGTACTGCACCGTGAGCTGGTCGGGAGAGCAATGTACGGCAAAGAGCCGAAATACTGTC   4020
                        _ClaI_
4021  ATTCATATGAGCAGGAATATCCGAT  4044  (SEQ ID NO:5)
```

FIGURE 8 page 3 of 3

TTTTGGTTTGCTGCNCGTTTGGGATAACTGCATAGAGAGCGGCCAAGTCGCTTGCGGTCG
      10   30   50

GTATCTCGAGTATATCGAAATCCATGTGGCCATTGACCTCTTCAAGCGCTCACGTTAACT
      70   90   110

ACCTGCTCTTTTTTGAGCACCAACATCCCAGGTTCGTCACAGTAAATCGTATCGTGATTA
      130   150   170

TTGCTAATCGTCAGTTTACCGCTCCGAAAGCAAACTANAAGTGAAACTGCTTACATAAAG
      190   210   230

ATTTTTGATGGTAACCTGCTGAGTCTGACTTTTAATTTGCTGCCGGGTATTTGTCAAAAG
      250   270   290

TGATTTTAATTTCTGTAAGTTATCTGCGGCAGGACGCTGATGACTATTACTTACAAAGGT
      310   330   350

TACATTTTCCATATTATCCCTTTGTTGAACTTATTTTAATGTTCCTTACTGGTATCCTAC
      370   390   410

TGAAAAAATCTGAGTTGTAAATGCTCTTTATTAGCGTGTGTTGGCAATGGTCTGATTGTT
      430   450   470

ACACCAAAAGAACCCAAATTTGGGTAATTTATCTACAGTAGTTTAAGCCCCAATGGGGAT
      490   510   530

GATGGTTCTTTTAATATGTGTTGAGACGCATTATACAGAATAAATTGATTTTATTTCTCA
      550   570   590

CTTTTCATTCTATTTTCATCAGGAATCCCTGTGTCCTGTGCGGTAATCTGCTGCTATCGA
      610   630   650
                                   prgH

GAACGACAGACATCGCTAACAGTATAT<u>ATG</u>GAAACATCAAAAGAGAAGACGATAACAAGC
                              M E T S K E K   T   I   T   S
      670   690   710

FIGURE 9 page 1 of 5

```
.CCAGGGCCATACATAGTTCGATTACTTAACAGCTCACTGAACGGCTGTGAGTTTCCATTG
 P  G  P  Y  I  V  R  L  L  N  S  S  L  N  G  C  E  F  P  L  -
      730                 750                    770
CTGACAGGCCGAACACTCTTTGTGGTAGGTCAGAGTGATGCGCTCACTGCTTCAGGTCAA
 L  T  G  R  T  L  F  V  V  G  Q  S  D  A  L  T  A  S  G  Q
      790                 810                    830

CTCCCTGATATACCTGCCGATAGCTTTTTTATCCCGCTGGACCATGGCGGAGTAAATTTT
 L  P  D  I  P  A  D  S  F  F  I  P  L  D  H  G  G  V  N  F
      850                 870                    890

GAAATCCAGGTGGATACGGATGCGACCGAAATTATACTCCATGAGCTGAAAGAAGGAAAT
 E  I  Q  V  D  T  D  A  T  E  I  I  L  H  E  L  K  E  G  N
      910                 930                    950

TCTGAATCTCGTTCGGTGCAATTAAATACGCCAATACAGGTCGGTGAATTGCTTATCCTG
 S  E  S  R  S  V  Q  L  N  T  P  I  Q  V  G  E  L  L  I  L
      970                 990                    1010

ATTCGCCCGGAAAGCGAGCCGTGGGTGCCCGAGCAGCCTGAGAAGTTAGAAACGTCTGCA
 I  R  P  E  S  E  P  W  V  P  E  Q  P  E  K  L  E  T  S  A
     1030                1050                   1070

AAAAAGAACGAGCCGCGTTTTAAAAACGGAATTGTAGCAGCACTGGCCGGGTTTTTTATA
 K  K  N  E  P  R  F  K  N  G  I  V  A  A  L  A  G  F  F  I
     1090                1110                   1130

TTGGGAATTGGGACTGTGGGGACGTTATGGATACTTAACTCGCCGCAGCGGCAGGCCGCA
 L  G  I  G  T  V  G  T  L  W  I  L  N  S  P  Q  R  Q  A  A
     1150                1170                   1190

GAGCTCGATTCGTTATTGGGGCAGGAGAAGGAGCGTTTTCAGGTGTTGCCAGGCCGGGAC
 E  L  D  S  L  L  G  Q  E  K  E  R  F  Q  V  L  P  G  R  D
     1210                1230                   1250

AAAATGCTCTATGTCGCTGCGCAAAATGAAAGAGATACGTTGTGGGCTCGTCAGGTTTTA
 K  M  L  Y  V  A  A  Q  N  E  R  D  T  L  W  A  R  Q  V  L
     1270                1290                   1310

GCGAGGGGCGATTATGATAAAAATGCGCGAGTGATTAACGAAAACGAAGAAAATAAGCGT
 A  R  G  D  Y  D  K  N  A  R  V  I  N  E  N  E  E  N  K  R  -
     1330                1350                   1370

ATCTCTATCTGGCTGGATACCTATTATCCGCAGCTGGCTTATTATCGGATTCATTTCGAT
 I  S  I  W  L  D  T  Y  Y  P  Q  L  A  Y  Y  R  I  H  F  D -
     1390                1410                   1430

GAGCCGCGTAAACCCGTTTTCTGGCTAAGCCGCCAGCGAAACACGATGAGCAAGAAAGAG
 E  P  R  K  P  V  F  W  L  S  R  Q  R  N  T  M  S  K  K  E
     1450                1470                   1490

CTCGAGGTGTTAAGTCAAAAGCTGAGAGCGCTAATGCCTTACGCGGATTCGGTTAACATC
 L  E  V  L  S  Q  K  L  R  A  L  M  P  Y  A  D  S  V  N  I
     1510                1530                   1550

ACGTTGATGGACGATGTTACCGCAGCAGGCCAGGCGGAAGCGGGGCTAAAACAGCAGGCG
 T  L  M  D  D  V  T  A  A  G  Q  A  E  A  G  L  K  Q  Q  A
```

FIGURE 9 page 2 of 5

```
            1570                    1590                    1610
TTACCTTATTCCCGCAGGAATCATAAGGGGGGCGTAACGTTTGTTATTCAGGGGGCGCTC
L   P   Y   S   R   R   N   H   K   G   G   V   T   F   V   I   Q   G   A   L
            1630                    1650                   1670

GATGATGTAGAAATACTCAGAGCCCGTCAATTTGTCGATAGCTATTACCGCACATGGGGA
D   D   V   E   I   L   R   A   R   Q   F   V   D   S   Y   Y   R   T   W   G
            1690                    1710                   1730

GGACGCTATGTGCAGTTTGCGATCGAATTAAAAGATGACTGGCTCAAGGGGCGCTCATTT
G   R   Y   V   Q   F   A   I   E   L   K   D   D   W   L   K   G   R   S   F
            1750                    1770                   1790

CAGTACGGGGCGGAAGGTTATATCAAAATGAGCCCAGGCCATTGGTATTTCCCAAGCCCA
Q   Y   G   A   E   G   Y   I   K   M   S   P   G   H   W   Y   F   P   S   P
            1810                    1830                      1850
                                       prgI
CTTTAATTTAACGTAAATAAGGAAGTCATTATGGCAACACCTTGGTCAGGCTATCTGGAT
L   ***  (SEQ ID NO: 11)                M   A   T   P   W   S   G   Y   L   D
            1870                    1890                   1910

GACGTCTCAGCAAAATTTGATACGGGCGTTGATAATCTACAAACGCAGGTAACAGAGGCG
D   V   S   A   K   F   D   T   G   V   D   N   L   Q   T   Q   V   T   E   A
            1930                    1950                   1970

CTGGATAAATTAGCAGCAAAACCCTCCGATCCGGCGCTACTGGCGGCGTATCAGAGTAAG
L   D   K   L   A   A   K   P   S   D   P   A   L   L   A   A   Y   Q   S   K
            1990                    2010                    2030

CTCTCGGAATATAACTTGTACCGTAACGCGCAATCGAACACGGTAAAAGTCTTTAAGGAT
L   S   E   Y   N   L   Y   R   N   A   Q   S   N   T   V   K   V   F   K   D
            2050                    2070                   2090
                                                           prgJ
ATTGATGCTGCCATTATTCAGAACTTCCGTTAATCAGTTATAAGGTGGATTATGTCGATT
I   D   A   A   I   I   Q   N   F   R   *  (SEQ ID NO:12)      M   S   I
            2110                    2130                   2150

GCAACTATTGTCCCTGAGAATGCCGTTATAGGGCAGGCGGTCAATATCAGGTCTATGGAA
A   T   I   V   P   E   N   A   V   I   G   Q   A   V   N   I   R   S   M   E
            2170                    2190                   2210

ACGGACATTGTCTCGCTGGATGACCGGCTACTCCAGGCTTTTTCTGGTTCGGCGATTGCC
T   D   I   V   S   L   D   D   R   L   L   Q   A   F   S   G   S   A   I   A
            2230                    2250                   2270

ACGGCTGTGGATAAACAGACGATTACCAACAGGATTGAGGACCCTAATCTGGTGACGGAT
T   A   V   D   K   Q   T   I   T   N   R   I   E   D   P   N   L   V   T   D
            2290                    2310                   2330

CCTAAAGAGCTGGCTATTTCGCAAGAGATGATTTCAGATTATAACCTGTATGTTTCTATG
P   K   E   L   A   I   S   Q   E   M   I   S   D   Y   N   L   Y   V   S   M
            2350                    2370                   2390
                                                       prgK
GTCAGTACCCTTACTCGTAAAGGAGTCGGGGCTGTTGAAACGCTATTACGCTCATGATTC
V   S   T   L   T   R   K   G   V   G   A   V   E   T   L   L   R   S   ***  (SEQ ID NO:13)
            2410                    2430                   2450   M   I   R
```

FIGURE 9 page 3 of 5

```
GTCGATATCTATATACTTTTCTGCTGGTAATGACCCTTGCCGGCTGTAAGGATAAGGATC
-  R  Y  L  Y  T  F  L  L  V  M  T  L  A  G  C  K  D  K  D  L
      2470              2490                 2510

TTTTAAAAGGACTGGACCAGGAACAGGCTAATGAGGTCATTGCCGTTCTGCAAATGCACA
   L  K  G  L  D  Q  E  Q  A  N  E  V  I  A  V  L  Q  M  H  N
      2530              2550                 2570

ATATAGAGGCGAATAAAAATTGATAGCGGAAAATTGGGCTATAGCATTACCGTTGCTGAGC
   I  E  A  N  K  I  D  S  G  K  L  G  Y  S  I  T  V  A  E  P
      2590              2610                 2630

CTGATTTTACCGCTGCGGTGTACTGGATTAAAACTTATCAGCTTCCTCCCCGGCCACGGG
   D  F  T  A  A  V  Y  W  I  K  T  Y  Q  L  P  P  R  P  R  V
      2650              2670                 2690

TGGAAATAGCGCAGATGTTCCCGGCGGATTCGCTGGTATCGTCTCCGCGAGCTGAAAAGG
   E  I  A  Q  M  F  P  A  D  S  L  V  S  S  P  R  A  E  K  A
      2710              2730                 2750

CCAGGTTATATTCGGCTATTGAACAGCGACTGGAACAGTCATTACAGACGATGGAGGGCG
   R  L  Y  S  A  I  E  Q  R  L  E  Q  S  L  Q  T  M  E  G  V
      2770              2790                 2810

TGCTCTCCGCCAGGGTCCATATTAGTTATGATATTGATGCTGGTGAAAATGGCCGCCCGC
   L  S  A  R  V  H  I  S  Y  D  I  D  A  G  E  N  G  R  P  P
      2830              2850                 2870

CAAAACCTGTTCATCTGTCGGCATTAGCCGTATATGAACGAGGTTCGCCGCTTGCGCATC
   K  P  V  H  L  S  A  L  A  V  Y  E  R  G  S  P  L  A  H  Q
     ·2890              2910                 2930

AGATCAGCGATATCAAGCGTTTCTTAAAGAATAGTTTTGCCGATGTGGATTATGACAACA
   I  S  D  I  K  R  F  L  K  N  S  F  A  D  V  D  Y  D  N  I
      2950              2970                 2990

TTTCTGTTGTGTTGTCAGAACGTTCTGATGCCCAATTACAGGCTCCCGGCACACCAGTAA
   S  V  V  L  S  E  R  S  D  A  Q  L  Q  A  P  G  T  P  V  K
      3010              3030                 3050

AACGTAATTCTTTTGCAACCAGTTGGATTGTTTTGATTATTTTGTTATCCGTGATGTCAG
   R  N  S  F  A  T  S  W  I  V  L  I  I  L  L  S  V  M  S  A
      3070              3090                 3110

CAGGCTTTGGCGTCTGGTATTACAAAAACCATTATGCCCGCAATAAGAAAGGCATAACGG
   G  F  G  V  W  Y  Y  K  N  H  Y  A  R  N  K  K  G  I  T  A
      3130              3150                 3170

CTGATGATAAGGCGAAATCGTCAAATGAATAGGCAGCCATTACCCATTATCTGGCAAAGA
   D  D  K  A  K  S  S  N  E  ***  (SEQ ID NO:14)
      3190              3210                 3230

TCATTTTTGATCCGTTATCGTATATCCATCCTCAGCGGTTGCAGATAGCGCCGGAAATGA
      3250              3270                 3290

TTGTCAGACCGCGCCACGCGAAATGAGTTAATACTGGCGGCATGGCGGCGGCTTAAGAAC
      3310              3330                 3350
```

FIGURE 9 page 4 of 5

GGAGAAAAGGAGTGTATTCAAAACTCACTGACGCAGCTGTGGCTGCTCAGTGGCGCCGAC...
    3370   3390   3410

TGCCGCAAGTAGCGTATTTACTAAACTGAGAGCCGATCTGGCAAGGCAGGGAGCCTTGCT
    3430   3450   3470

TGGCCTAGCCGGATTGGGCGAAATGAGTTAATACTGGCGGCATGGCGGCTTGCCAT   (SEQ ID NO:1(
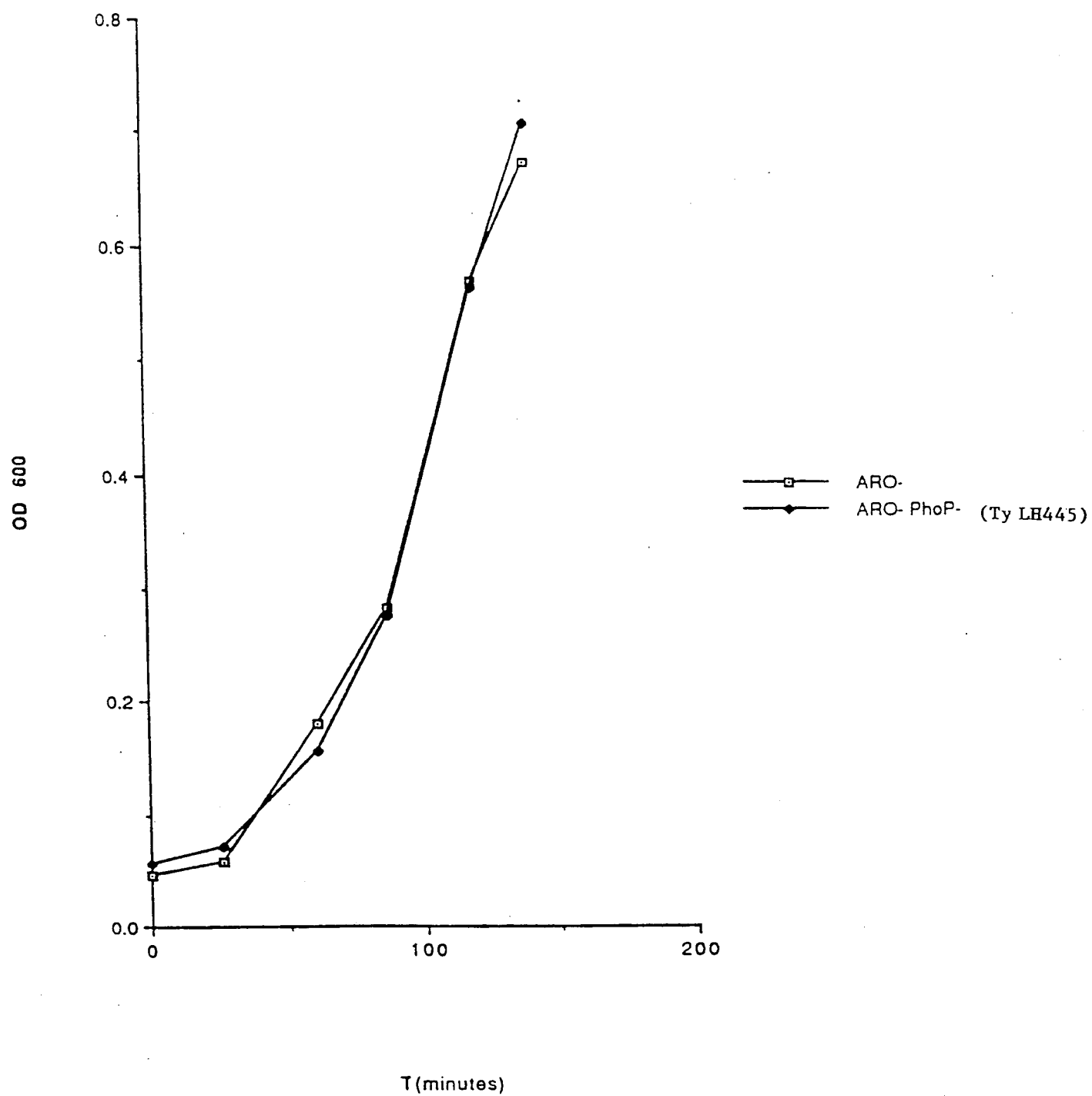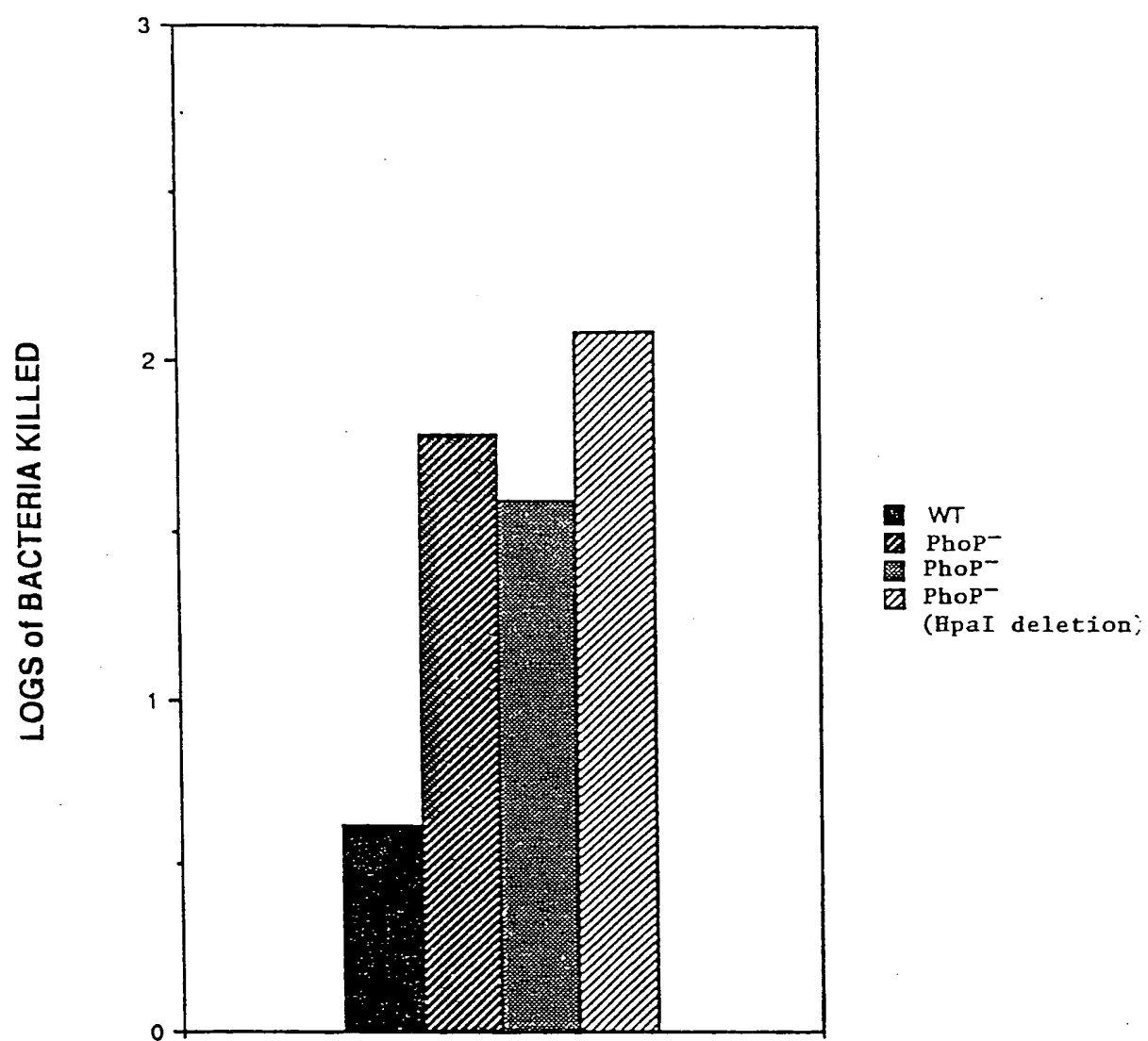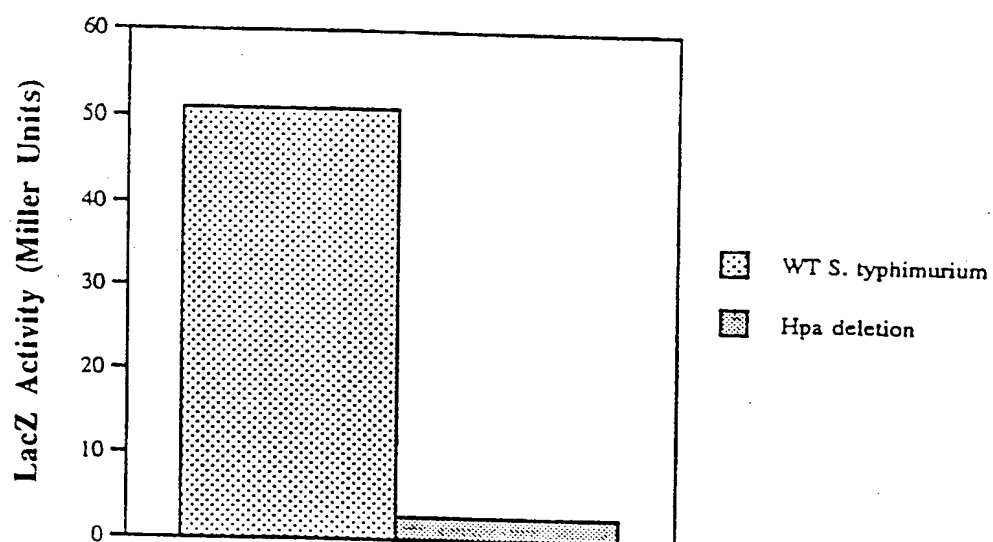   3490   3510   3530

FIGURE 9 page 5 of 5

FIGURE 10
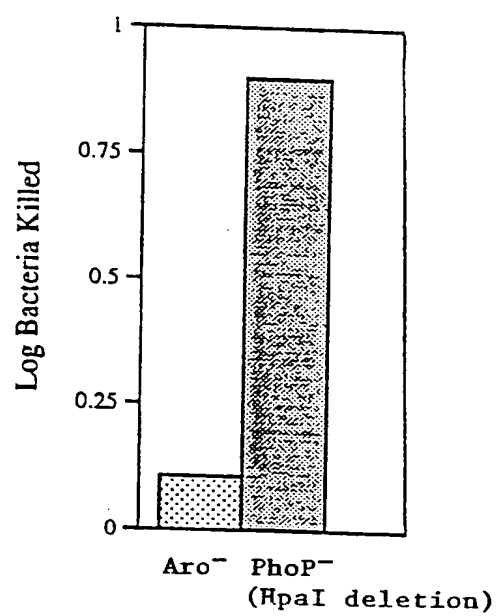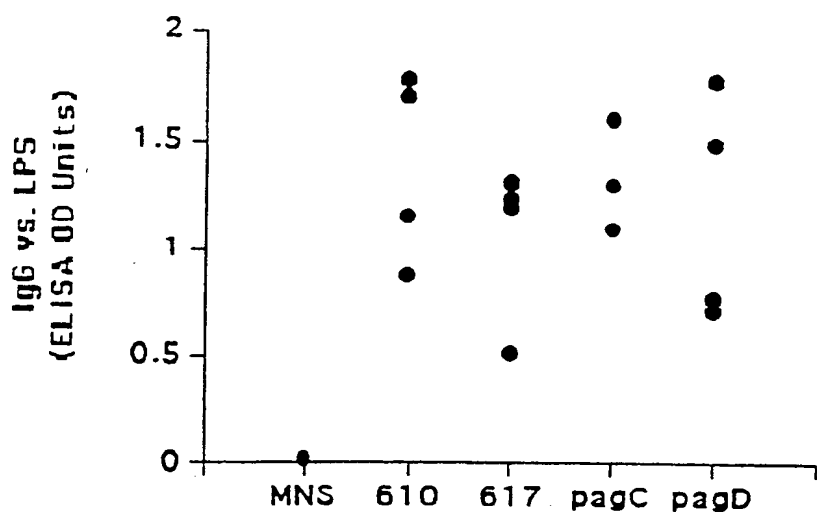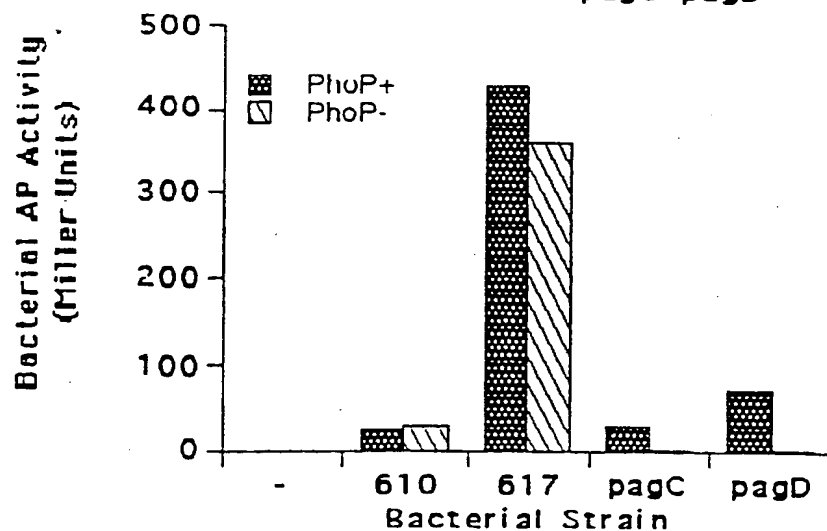
FIGURE 11

FIGURE 12

FIGURE 13
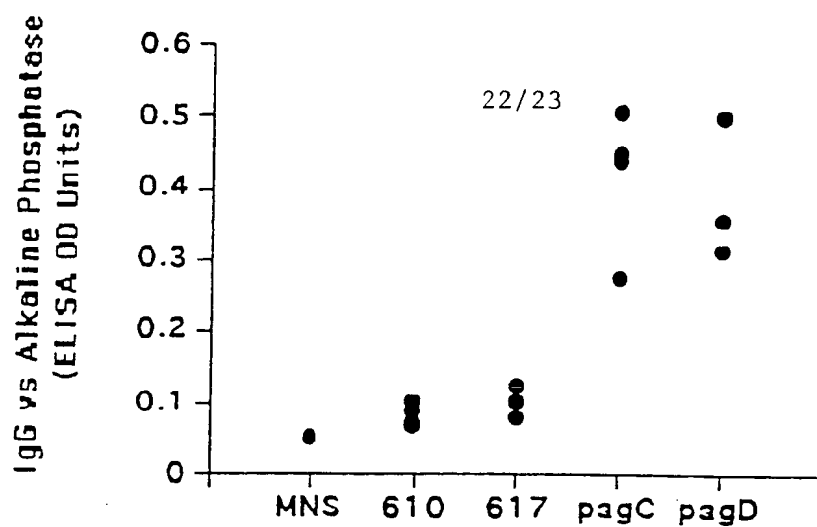
FIG. 14 C.

FIG. 14 B.

FIG. 14 A.

CATAACAACTCCTTAATACTACTTATTATTTACGGTGTGTTTAAACACCT 50

GCAGTACCGATCCGGCATTCAGTTATCGCCACTATGCCGAATCGACAAAA 100

CCACGAATAATTCACCGCTATCGCTCCTGATGTGTTACTTCCTGAAAGA 150

TATTTTTACTACCGAAGCACTCTATCGCTCATTTAGGTAACCGGTTCTAC 200

AATGTCATCTAACTTTTATAGATTTGAATGCTAATTTTTCTCACGCATAT 250

ATATTTAACAGAAACCATAAAGTGTTTAGCCACTATAGAACAACAAATCA 300

CCCATGCAACATTTTGATATTTAAAGAGAAATCTCACAACCACATTAAG 350

AAACTTGACACCGTTCGGCTAAAAACATGTCATTAAGCAAACTCGCCATA 400

TAATCAGAACATATCGCATTGTGCTTCACAGTCCTCACGTGACGCTCCAT 450

CCGCAATACGGTTATATGCCATCGCAGGCGCTGTAATCATATTCACGATG 500

ATGCTTAGCACGCTTTATTCCCGCTCCGATTTAATCTTTTAATATATCTA 550

TCAGTTACAACATTTCTTGTTATATTATAAGAATAGAATCAACACCACAA 600

TTCCAACATAAATATCACCTGTGTTTAGAGAGAATTTACATTCCAAAAAA 650

ATAATAACTAACGCAAATATTGAACACGCGATAAAAAGTCTATTTCGCT 700

ATAAAACCCATTATTATTAAGAGTGGTTAACTCTTCGTTGAATAAAAAT 750

GTCAATGACGTTCCATAATTCAGGAGATGAACTTCACAAGTCATTATATA 800

TAACAGGAGGTGCTATG 817   (SEQ ID NO:15)

FIGURE 15

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12N 15/00, 7/00; C12Q 1/00; A61K 39/02, 39/40

US CL :424/93A, 241.1, 258.1; 435/252.3, 252.8, 320.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/93A, 241.1, 258.1; 435/252.3, 252.8, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG, MEDLINE, CAS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO, A, 92/11361 (MILLER ET AL.) 09 JULY 1992, see entire document.	1-196
Y	Vaccine, Vol. 11, No. 2, issued 1993, Miller et al, "The PhoP Virulence Regulon and Live Oral Salmonella Vaccines", pages 122-125, see entire article.	1-196
Y	Molecular Microbiology, Vol. 5, No. 9, issued 1991, Miller, "PhoP/PhoQ: Macrophage-Specific Modulators of *Salmonella* Virulence?", pages 2073-2078, see Abstract on page 2073, and Table 1, page 2074, page 2078.	1-104, 115-123

X	Further documents are listed in the continuation of Box C.		See patent family annex.

	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 SEPTEMBER 1994	0 7 OCT 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer H. F. SIDBERRY
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Research Microbiology, Vol. 141, issued 1990, Miller et al, "*Salmonella* Vaccines With Mutations In The *phoP* Virulence Regulon", pages 817-821, see pages 817, 819 and 820.	1-196

THIS PAGE BLANK (USPTO)